



Hochschule Karlsruhe – Technik und Wirtschaft
Fakultät für Elektro- und Informationstechnik
Studiengang Elektrotechnik – Sensorik

BACHELORTHESIS

von

Sebastian Bek

Videobasierte Dichte- und Flussschätzung in Menschenmassen

Arbeitsplatz:	Fraunhofer IOSB
Betreuer am Arbeitsplatz:	Dr.-Ing. Eduardo Monari
Referent:	Prof. Dr.-Ing. Ulrich Grünhaupt
Korreferent:	Prof. Dr. Thomas Westermann
Abgabetermin:	03.03.2016
Bachelorthesis Nr.:	255 des Studienganges Elektrotechnik – Sensorik
Bearbeitet in der Zeit vom	03.12.2015 bis 03.03.2016

Erklärung der Selbstständigkeit

Ich versichere hiermit, dass ich die vorliegende Bachelorthesis selbstständig erarbeitet und dabei keine anderen als die angegebenen Quellen verwendet habe.

Karlsruhe, den 2. April 2019

Sebastian Bek

Abstract

In den letzten Jahren erreichen uns immer häufiger Nachrichten über große Veranstaltungen, bei denen sich aufgrund zu hoher Besucherzahlen in Kombination mit unzureichender Planung Tumulte und Paniken, häufig mit Verletzten, ausbilden. Solche Großveranstaltungen werden immer beliebter und sind fester Bestandteil der menschlichen Freizeitaktivitäten geworden, weswegen Einschränkungen dieser unwahrscheinlich sind. Die Tendenz zeigt, dass mit steigenden Besucherzahlen auch die Gefahr für Paniken und Unruhen steigt, weswegen videogestütztes Monitoring hier sinnvoll sein kann. Derzeit erfolgt das Monitoring von Veranstaltungen noch durch menschliche Betrachter, die oft unaufmerksam sind und meist nur subjektiv beurteilen können, wie hoch die Personenströme und -dichten wirklich sind. Zukünftig wird sich aber das automatisierte Monitoring durchsetzen, mit dem es möglich ist, ohne menschliches Eingreifen, Besucherzahlen und Personenströme quantitativ und relativ genau zu messen. Die automatische Erfassung und Analyse von Menschenströmen stellt jedoch zum heutigen Zeitpunkt noch eine technische Herausforderung dar.

In den ersten Jahren seines Lebens lernt der Mensch Formen, Strukturen und Kontouren richtig zu klassifizieren und aufgrund von Erfahrungswerten richtig einzuordnen. Mit diesen Erfahrungswerten, die sich über eine lange Entwicklung hin gebildet haben, kann der Mensch, mit einer sehr hohen Trefferquote, Personen als solche von anderen Strukturen in Bildern unterscheiden. Rechner und entsprechende Algorithmen sind hinsichtlich ihrer Erkennungsleistung noch lange nicht mit dem Menschen vergleichbar. Insbesondere bei niedriger Auflösung und starken Verdeckungsartefakten in der Szene, ist eine automatische Detektion von Personen mit den bekannten Verfahren bisher kaum möglich.

Im Gegenzug besitzt der Rechner, im Falle einer möglichen Detektion von Objekten, signifikante Vorteile bei der Analyse großer Datenmengen und bei der quantitativen Abschätzung von Messgrößen gegenüber der visuellen Auswertung. Um die Erkennungsleistung eines automatischen Systems, trotz der schwierigen Rahmenbedingungen bei der Beobachtung von Menschenmassen (niedrige Auflösung und Verdeckungen) auf ein einsetzbares Niveau zu steigern, werden im Rahmen dieser Arbeit Methoden basierend auf „indirekten Merkmalen“, wie beispielsweise Bewegungsmerkmalen, untersucht, um auf Basis dieser Informationen über die lokale Dichte und den lokalen Personenfluss innerhalb der Menschenmenge zu erschließen. Diese Dichte- und Flussinformationen sollen klassifiziert, quantifiziert und veranschaulicht werden, um etwaige Folgemaßnahmen bei kritischen Situationen zu erschließen.

Kurzfristige Folgemaßnahmen könnten z. B. Evakuierung und Absperrung bestimmter Gebiete oder Eingriffe in Unruhen sein. Langfristige Folgemaßnahmen können z. B. eine verbesserte Planung der Infrastruktur der Veranstaltung oder die Wahl eines größeren Veranstaltungsgeländes sein. Es soll ein Verfahren zur Erkennung und Messung der Bewegungs- und Gruppierungsmuster entwickelt und hinsichtlich seiner Performanz bewertet werden, um zu zeigen, dass es, auch unter Verwendung von niedrig aufgelösten Videodaten, für die Analyse von Dichte- und Flussinformationen in dicht gedrängten Personengruppen geeignet ist.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Aufgabenstellung und Ziele	2
1.3	Stand der Technik	3
1.3.1	direkter Ansatz	3
1.3.2	indirekter Ansatz	4
2	Grundlagen	7
2.1	Harris Corner Detector	7
2.2	Corner Tracker	10
2.3	Verschiebungsvektoren	12
2.4	Erfassung von Trajektorien	13
3	Systemarchitektur	17
3.1	Beschreibung des Messsystems	17
3.2	Signalverarbeitung	19
3.2.1	Trajektorien	19
3.2.2	Personenzahl	21
3.2.3	Personenfluss	22
3.2.4	Dynamik	23
3.2.5	Dichtefaktor	24
3.2.6	Darstellung der Messgrößen - Aufbereitung	24
3.3	Konfigurierung	26
3.3.1	Patch-Map Dimension	26
3.3.2	Geschwindigkeitsschwelle	27
3.3.3	Maximale Verweilzeit	27
3.3.4	Mittelungsdistanz	28
3.3.5	Suchradius	28
3.3.6	Abstands-Radius	30
3.3.7	Ortsbezogener Ausgleich	30
3.3.8	Distanzschätzung	32
3.3.9	Filtereinstellungen	33

4	Evaluation	35
4.1	Testdaten	35
4.1.1	3D-Animation: AGORASET/Dispersion [?] [?]	35
4.1.2	Videosequenz: UCF_CrowdsDataset/Business-Getümmel [?]	36
4.1.3	Videosequenz: UCF_CrowdsDataset/Marathon [?]	36
4.1.4	Videosequenz: Hamburger Hafenfest 2014	37
4.2	Testergebnisse	41
4.2.1	3D-Animation: AGORASET/Dispersion [?] [?]	41
4.2.2	Videosequenz: UCF_CrowdsDataset/Business-Getümmel [?]	46
4.2.3	Videosequenz: UCF_CrowdsDataset/Marathon [?]	50
4.2.4	Videosequenz: Hamburger Hafenfest 2014	50
5	Fazit	53
	Abbildungsverzeichnis	55

Einleitung

Diese Arbeit befasst sich mit der algorithmischen Detektion und Messung von Bewegungs- und Gruppierungsmustern von Personen, wobei Trajektorien (Pfade) als Grundlage für die darauf aufsetzende Dichte- und Flussschätzung verwendet werden. Besonders intensiv werden dabei die Personenzählung (in Personen/Bildausschnitt), die Berechnung einer mittleren Strömung (in Personen/Frame) und die Ableitung eines Dichtefaktors (in % einer Maximaldichte) in einem bestimmten Bildausschnitt behandelt. Das Verfahren soll in Kombination mit Anderen in Form einer integrierten Software zur komfortablen Analyse von Gruppierungs- und Bewegungsmustern und zur Einrichtung von Schnellwarnsystemen eingesetzt werden.

Der nachfolgende Abschnitt (1.1) befasst sich mit der Motivation für die automatisierte Videoauswertung. Er basiert auf der Dissertation „Automatische Erfassung präziser Trajektorien in Personenströmen hoher Dichte“ von Maik Boltes [?] und dem Artikel „A Method for Counting Moving People in Video Surveillance Videos“ von Donatello Conte et al. [?]. Anschließend werden die Ziele dieser Arbeit (s. Abschnitt 1.2) und im folgenden Abschnitt 1.3 der aktuelle Stand der Technik behandelt.

1.1 Motivation

Personenströme und -anordnungen im Alltag sind für uns gewöhnliche Abläufe, die aber dennoch interessante Erkenntnisse über die Selbstorganisation und die Gewohnheiten der Menschen liefern können. Zudem wäre es mit Sicherheit von Vorteil ein tiefgreifendes Verständnis der bisher nur grob erforschten Dynamik von Personenströmen zu gewinnen, um Beobachtungen zu machen, die in sicherheitskritischen und ökonomischen Anwendungsgebieten hilfreich sein können.

Die Erfassung dicht gedrängter Personengruppen mit Kameras und Bildauswertungssoftware erweist sich dabei als besonders zielführend, da analysierte Szenen vom Computer direkt ausgewertet und die angezeigten Informationen aufbereitet werden können. An das Personal werden dabei geringere Anforderungen an Aufmerksamkeit und Organisation gestellt. Es können außerdem Messdaten aufgenommen und quantitative Aussagen über das Verhalten von Personengruppen gemacht werden.

Gestützte Monitoringssysteme sind vor allem an Orten, die von vielen Menschen gleichzeitig besucht werden (z. B. Großveranstaltungen oder Fußgängerzonen) sinnvoll. Denn mit höheren Besucherzahlen nimmt möglicherweise auch das Risiko für Schadensereignisse oder Kriminalität zu, weil die Zahl an Personen, die bereit sind Verbrechen zu begehen und Tumulte zu verursachen, mit der Besucherzahl steigt. Solche Großveranstaltungen bergen zudem die Gefahr für Massenpaniken, weswegen als Grundsatz für eine Veranstaltungsplanung Flucht- und Rettungswege ausnahmslos freigehalten werden sollten. Langfristig kann mit flächigem videogestütztem Monitoring die Planung verbessert werden, indem z. B. kritische Gebiete entlastet werden (breitere Tore/Türen, Fluchtwege) oder die Zeitplanung für die öffentlichen Verkehrsmittel überdacht wird. Für kurzfristige Reaktionen kann mit der integrierten Software ein Schnellwarnsystem, in Form eines Kameranetzwerkes, am Veranstaltungsort installiert werden, das schnell ortsbezogene Notrufe tätigen und exakte Positionen von besonders kritischen Gebieten des Veranstaltungsgeländes übermitteln kann. Kritische Gebiete sind zum Beispiel Staugebiete, also Orte, an denen die Personenzahl hoch ist und eine geringe Dynamik herrscht. Ein solches Schnellwarnsystem sollte kritische Situationen schnell erkennen und melden können, weswegen als Voraussetzung die Extraktion der Laufwege auch bei hohen Personendichten (, die kritischen Situationen entsprechen,) verlässlich anwendbar sein muss. Außerdem ist es möglich, durch eine Kalibrierung der Kameras, georegistrierte Messwerte zu erhalten, die beispielsweise auf einer Karte oder Aufnahme des Veranstaltungsgeländes dargestellt werden können. Dies ermöglicht eine einfache visuelle Auswertung der Daten, die geringe Anforderungen an die Aufmerksamkeit der Nutzer stellt. Die Daten können entweder in Echtzeit oder im Nachhinein zur verbesserten Planung analysiert werden. Zusätzlich kann das aufgenommene Beweismaterial zu Rate gezogen werden, wenn Unklarheiten über die Ursache für den sicherheitskritischen Fall bestehen. Dies sollte aber ohne Identifizierung von Einzelnen erfolgen.

1.2 Aufgabenstellung und Ziele

In dieser Arbeit wird die Untersuchung und Entwicklung eines Videoanalyseverfahrens zur automatischen Extraktion von Trajektorien (Pfaden) bewegter Personen behandelt. Das Verfahren soll im Rahmen eines Videoanalysesystems, das für die Realisierung eines Schnellwarnsystems und für die komfortable Analyse von Bewegungs- und Gruppierungsmustern (s. Abschnitt 1.1) zur Langzeitplanung von Veranstaltungen geeignet ist, entwickelt und dokumentiert werden.

Es sollen Laufwege bewegter Personen, deren Ecken zuvor algorithmisch detektiert wurden, als Bildkoordinaten in Listen eingetragen und verwaltet werden. Solche Laufwege werden als Trajektorien bezeichnet. Dabei sollen nur bewegte Eckpunkte verwendet werden, weswegen gänzlich stillstehende Personen (zunächst) nicht erfasst werden. Weiterhin können aus den gespeicherten Positionen der Trajektorien, innerhalb von bestimmten Bildausschnitten, Messgrößen wie Personenfluss, Personenzahl und Dichtefaktoren abgeleitet werden. Das Verfahren sollte zuverlässig sein und keine auffälligen, kritischen Situationen übersehen. Es soll erprobt und mithilfe von Grundwahrheiten bewertet werden, um eine Aussage über die Wahrheitstreue der Messgrößen treffen zu können.

1.3 Stand der Technik

Die Idee, automatische Videoanalyseverfahren zur Erkennung von Mustern und Objekten einzusetzen, ist keine neue Vorstellung. Wegen der zunächst nur wenig fortgeschrittenen Technik, setzte sich diese Idee jedoch erst in den 1980er Jahren durch. Heute gibt es vielseitige Ansätze, um diese Aufgaben zuverlässig durchzuführen. Dennoch haben diese Verfahren ihre Grenzen und es existiert ein weitreichendes Verbesserungspotenzial.

Wie in Abschnitt 1.2 beschrieben, wird in dieser Arbeit die algorithmische Detektion und Messung von Dichte- und Flussinformationen in Personengruppen behandelt. Um solche Gruppierungs- und Bewegungsmuster erfassen und messen zu können, müssen die Personen zunächst quantitativ erfasst werden. Zur Zählung von Personen/Objekten existieren, nach aktuellem Stand, zwei grundlegende Ansätze. Zunächst muss, in beiden Fällen, eine Extraktion einer ROI (region of interest - z. B. Punkte, Flächen) zur groben Ausfilterung von uninteressanten Bereichen, an denen sich gerade keine Personen befinden können, vorgenommen werden. Dabei werden uninteressante, statische Bereiche im Bild ausgefiltert, während die interessanten Bereiche hinsichtlich Personendetektion, Dichte- und Flussschätzung weiterverarbeitet werden. Nachfolgend werden die beiden Ansätze beschrieben und mit Beispielen ausgeführt.

1.3.1 direkter Ansatz

1. Segmentierung von VG-Bereichen: nach Möglichkeit mit beinhalteten Personen
2. Detektion und individuelle Separation der Personen
3. Zählung der separierten Vordergrund-Regionen

Im Fall des direkten Ansatzes, werden in Kombination zur Extraktion einer ROI in VG-Regionen Muster und Merkmale gesucht, die Menschen zugeordnet werden sollen, um diese einzeln zu separieren.

Nachfolgend werden einige Beispiele für den direkten Ansatz aufgeführt:

Objektklassifizierung:

- Vergleich der Form: Kanten von Vordergrund-Regionen werden mit Modellen (Rechtecke, Ellipsen o.ä.), die entweder ganze Menschen oder Körperteile modellieren, verglichen (s. [2]).
- Vergleich der Kantenstatistik: Anzahl, Ausrichtung und Form der Kanten können Menschen beschreiben, wobei Statistiken/Histogramme über die Häufigkeit von Kantenrichtungen erstellt und gespeichert werden, falls sich ein/kein Mensch im Bildausschnitt befindet (s. HOG = histogram of oriented gradients). Anschließend kann, über einen Vergleich der Kantenstatistik des analysierten Bildausschnitts mit der gespeicherten Kantenstatistik, entschieden werden, ob ein Mensch in einem Bildausschnitt befindlich ist. Dies wird von einem lernenden Algorithmus (SVM: support vector machine) durchgeführt. Zusätzlich kann hier, wie in [2] beschrieben, bestimmt werden, ob sich gerade mehr (komplexere Kanten) oder weniger Menschen (einfache Kanten) in einem größeren Bildausschnitt aufhalten.
- Vergleich von aufgenommenen 3D-Modellen mit Standardmodellen: Nach [2] können Menschen erkannt werden, indem ein 3D-Modell, das mit einer 3D-Kamera oder einem Sensor-Array erstellt wurde, mit Standardmodellen für Menschen verglichen wird.
- Gruppieren von extrahierten Merkmalen („Features“ - z. B. Punkten) nach ihren Bewegungscharakteristika: Punkte, die sich nahezu gleich schnell in eine Richtung bewegen, gehören mit hoher Wahrscheinlichkeit nur zu einer Person (s. [2]).

Anschließend wird, über ein Labeling-Verfahren (z. B. ein sog. „Point Clustering“), eine Vergabe von Identifikationsnummern durchgeführt, wobei die einzelnen (als Vordergrund-Regionen) separierten Personen mit IDs versehen und gezählt werden können.

Beim nachfolgenden indirekten Ansatz wird eine solche Objektklassifizierung nicht durchgeführt, sondern direkt mit der extrahierten ROI gearbeitet.

1.3.2 indirekter Ansatz

1. Detektion und Extraktion von lokalen Merkmalen (segmentierungslos - z. B. Punkte/Pixelpositionen), die beispielsweise durch extrahierte Eckpunkte (s. „Harris Corner Detector“ [2]) gegeben sind und in Kombination mit einer Bewegungsschätzung (optischer Fluss) als lokale Bewegungsmerkmale verwendet werden können.
2. Ableitung der Personenzahl aus der Zahl der lokalen, extrahierten Merkmale, unabhängig von einer separaten Detektion von Personen

Dieser Ansatz gilt als robuster, weil die vereinzelte Separation der Personen, besonders in überfüllten Gebieten, ein zu großes Problem darstellt und bisher nicht zuverlässig gelöst wurde.

Nachfolgend werden einige Beispiele für den indirekten Ansatz aufgelistet:

- Zählen der Menge an nicht-statischen Pixeln (s. [?])
- Zählen der Menge an nicht-statischen Eckpunkten, extrahiert von einem Corner Detector (z. B. Harris Corner Detector, wie in [?] beschrieben)
- Erfassen der Größen und Häufigkeiten von VG-Bereichen (vgl. [?])
- Erfassen der fraktalen Dimension von VG-Bereichen (vgl. [?]):

$$D = \frac{\log(\text{Zahl der selbstähnlichen Bestandteile})}{\log(\text{Abbildungsmaßstab})}$$

In dieser Arbeit wird eine Ausführung des indirekten Ansatzes behandelt. Dabei kommt der Harris Corner Detector, wie in [?] beschrieben, zum Einsatz, um Eckpunkte/Merkmale zu extrahieren und vom Hintergrund zu trennen. Ein Corner Tracker erfasst die relative Bewegung (den optischen Fluss) dieser Merkmale entlang der Bildzeilen und -spalten. Die Anzahl an bewegten Personen wird in ein Verhältnis mit der Anzahl an sich bewegenden, extrahierten Eckpunkten gesetzt und daraus werden weitere Größen wie Personenfluss und ein Dichtefaktor abgeleitet.

Wichtige Verfahren, in Bezug auf diese Arbeit, sind somit zunächst der Harris Corner Detector, gekoppelt an einen Corner Tracker, der die Bewegung (den optischen Fluss) erfasst, sowie das Verfahren zur Erfassung von Trajektorien (Pfaden) der extrahierten Eckpunkte. Diese Grundlagen werden im Kapitel 2 näher erläutert. Diese Arbeit soll zeigen, dass das entwickelte Verfahren, auch mit den niedrig aufgelösten Daten z. B. von CCTV-Überwachungskameras, in Personengruppen hoher Dichte zuverlässig angewandt werden kann.

Grundlagen

Dieses Kapitel befasst sich mit den Basismethoden, welche in den untersuchten Verfahren eingesetzt werden. Zunächst wird in Abschnitt 2.1 das zugrundeliegende Verfahren zur Extraktion lokaler Bildmerkmale beschrieben, das auf dem Harris Corner Detector [?] basiert. Anschließend wird im Abschnitt 2.2 der verwendete Corner Tracker behandelt, welcher zu den lokalen Bildmerkmalen Bewegungsinformationen (vgl. Abschnitt 2.3) hinzufügt (motion vectors - Bewegungsvektoren). Im letzten Abschnitt 2.4 wird das verwendete Verfahren zur Erfassung von Trajektorien (Pfaden) der lokalen Bildmerkmale, und dadurch indirekt der Personen (-gruppen), näher beschrieben.

2.1 Harris Corner Detector

Der „Harris Corner Detector“ [?] ist ein Algorithmus, der verschiedene Bildausschnitte (Patches) einer Bildmatrix I nach Ecken (engl.: Corners) durchsucht und diese in Form der Position des Center Pixels extrahiert, falls sich eine Ecke im Bildausschnitt befindet. Eine grafische Beschreibung der Idee des Harris Corner Detector wird auf der folgenden Seite dargestellt. Dabei werden unterschiedliche Patches (s. Abbildung 2.1 - blau) mit unterschiedlichen Situationen betrachtet:

Abbildung 2.1: Betrachtung der Gradientenbeiträge verschiedener Patches

Diese anschauliche Beschreibung der Idee des Harris Corner Detectors zeigt, dass in der Umgebung einer Ecke immer viele hohe Intensitätsunterschiede (Gradientenbeiträge) in verschiedene Richtungen vorliegen. In Abbildung 2.2 werden Verteilungen der Gradienten-

beiträge in x- und y-Richtung für die verschiedenen Fälle in einem Bildausschnitt betrachtet:



Abbildung 2.2: Verteilungen von Gradientenbeiträgen I_x, I_y [?]]

Dabei wird klar, dass je nach Situation eine andere Verteilung der Daten vorliegt und zwischen allen drei Fällen klar unterschieden werden kann. Dafür muss aber unterschieden werden können, ob die Varianz der Verteilung mit Bezug zum Ursprung generell klein (im Fall einer Fläche), groß in eine Richtung (im Fall einer Kante) oder groß in mindestens 2 Richtungen (im Fall einer Ecke) ist. Dazu wird eine Kovarianzmatrix (auch bekannt als Strukturtensor) M für die Verteilung der Gradientenbeiträge I_x, I_y hergeleitet (Herleitung siehe [?]), die die Varianz der Daten mit Bezug zum Ursprungswert ($I_x = 0, I_y = 0$) im aktuellen Bildausschnitt mit Center Pixel x, y beschreibt:

$$M = \sum_{u,v} w(u,v) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} = \begin{bmatrix} a = \sum_{u,v} w(u,v) I_x^2 & b = \sum_{u,v} w(u,v) I_x I_y \\ b = \sum_{u,v} w(u,v) I_x I_y & c = \sum_{u,v} w(u,v) I_y^2 \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \text{var}(I_x) & \text{covar}(I_x, I_y) \\ \text{covar}(I_x, I_y) & \text{var}(I_y) \end{bmatrix} \quad (2.1)$$

mit $u, v \in W(u, v)$, $I_x = I_x(x + u, y + v)$ und $I_y = I_y(x + u, y + v)$

I_x und I_y werden wegen der kompakteren Schreibweise abgekürzt. $w(u, v)$ ist hier eine Gewichtungsfunktion wie die Rechteck- oder Gauss-Funktion. $W(u, v)$ beschreibt den zulässigen Wertebereich, innerhalb dem u und v liegen müssen, um den definierten Bildausschnitt nicht zu überschreiten. Für die Eigenwerte der Matrix M gilt:

$$\lambda_{1/2} = \frac{a + c \pm \sqrt{(a - c)^2 + 4b^2}}{2} \quad (2.2)$$

Im Fall einer Kovarianzmatrix geben die Eigenvektoren dieser die 2 orthogonalen Hauptrichtungskomponenten der Varianz der Daten an. Die Eigenwerte λ_1 und λ_2 beschreiben die Länge der größten Eigenvektoren und damit die Varianz der Verteilung von I_x, I_y in Richtung der Eigenvektoren. Damit ergeben sich im Falle einer:

- flachen Region (s. Abbildung 2.2 - flat): 2 kleine Eigenwerte der Matrix B (keine Varianz) $\Rightarrow \lambda_1 \sim \lambda_2 \approx 0$
- Kantenregion (s. Abbildung 2.2 - edge): 1 großer, 1 kleiner Eigenwert der Matrix B (große Varianz in eine Richtung) $\Rightarrow \lambda_1 \gg \lambda_2$ \vee $\lambda_2 \gg \lambda_1$
- Eckregion (s. Abbildung 2.2 - corner): 2 große Eigenwerte der Matrix B (große Varianz in mindestens 2 Richtungen) $\Rightarrow \lambda_1 \sim \lambda_2$

Um in allen Fällen verschiedene Ergebnisse zu erhalten, wird eine sogenannte Corner Response an jedem Pixel definiert, die große, positive Werte für Ecken, große Negative für Kanten und kleine Werte für flache Regionen annimmt:

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (2.3)$$

k ist ein empirisch festgelegter Parameter, dessen Wertebereich zwischen 0,04 und 0,06 liegt. Um den Rechenaufwand, der durch die Berechnung der Eigenwerte entsteht (Quadratwurzeln sind rechenintensiv) zu verringern, wird die Corner Response oft so definiert:

$$R = \det B - k(\text{spur } B)^2 \quad (2.4)$$

mit: $\det B = \lambda_1 \lambda_2$ und $\text{spur } B = \lambda_1 + \lambda_2$

2.2 Corner Tracker

Als Grundlage der Arbeit liegt eine Bibliothek „OCV_CornerTracker“ vor, in der im Rahmen eines sogenannten Corner Trackers der Harris Corner Detector bereits implementiert ist. Das Blockschaltbild dieser Bibliothek wird in Abbildung 2.3 dargestellt.

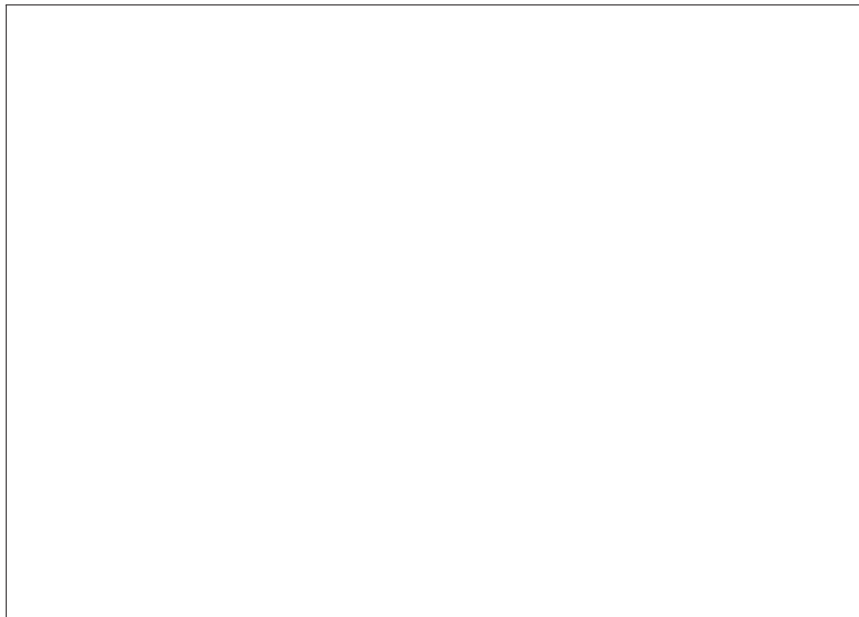


Abbildung 2.3: Blockschaltbild: OCV_CornerTracker

Funktion:

Das hier eingesetzte Verfahren zum Tracking von Harris Corner Features (Merkmale) berechnet zunächst für jeden Pixel im vorherigen (n-1.) Eingangsbild den sogenannten optischen Fluss [?]. Dabei wird ein Intensitätsprofil aus der Umgebung eines Pixels erstellt. Dieses Profil wird in der Umgebung des Pixels im aktuellen (n.) Eingangsbild gesucht.

Die relative Verschiebung des Mittelpunkts im Profil im Vektorraum, zwischen beiden Eingangsbildern, gibt den optischen Fluss des Pixels in Zeilen- und Spaltenrichtung an.

In Abbildung 2.4 werden beispielhaft solche Verschiebungsvektoren in einem Videobild in blau eingezeichnet. Anschließend wird zu den jeweiligen Pixelpositionen, die im vorherigen (n-1.) Eingangsbild als Harris Corners klassifiziert wurden, der zugehörige optische Fluss als Bewegungsschätzung hinzugefügt. An Positionen, an denen die Merkmale keine Bewegungen aufweisen oder keine Merkmale extrahiert wurden, werden die Verschiebungsvektoren gleich 0 gesetzt. Zusammenfassend enthalten die generierten Verschiebungsvektoren ausschließlich pixelbezogene Bewegungen der extrahierten Merkmale des Harris Corner Detectors vom letzten (n-1.) zum aktuellen (n.) Frame.



Abbildung 2.4: optischer Fluss von detektierten Ecken an Marathonläufern [?]]

2.3 Verschiebungsvektoren

Die Verschiebungsvektoren, die vom Corner Tracker (s. Abschnitt 2.2) ausgegeben werden, beinhalten die relativen Bewegungen der extrahierten Merkmale in Zeilen- und Spaltenrichtung am Ort des Ursprungs. Aus den beiden Vektorkomponenten können für jede detektierte Position im Bild Vektorbeträge berechnet werden:

$$|\vec{a}| = \sqrt{DX^2 + DY^2} \quad [\text{px}] \quad (2.5)$$

Weil diese Vektorlänge sich im Verfahren auf die Bewegung eines Merkmals über einen Frame hinweg bezieht, ist diese Angabe eine Merkmalsgeschwindigkeit:

$$v = \frac{|\vec{a}|}{\text{lframe}} = \frac{\sqrt{DX^2 + DY^2}}{\text{lframe}} \quad \left[\frac{\text{px}}{\text{frame}} \right] \quad (2.6)$$

Kennt man den genäherten Abstand $d(x,y)$ zwischen den Pixeln im betreffenden Bildausschnitt als abbildende Funktion (z. B. in Form einer Karte mit Einträgen in $[\frac{m}{px}]$) und zusätzlich die Framerate FR (in [fps]) erhält man die reale Merkmalsgeschwindigkeit genähert in SI-Einheiten:

$$v_{SI}(x,y) \approx v(x,y) \cdot d(x,y) \cdot FR \quad [\frac{m}{s}] \quad (2.7)$$

2.4 Erfassung von Trajektorien

In diesem Abschnitt wird das Verfahren zur Erfassung von Trajektorien (Pfaden) der Personen beschrieben. Die Funktion dieses Verfahrens wird in Abbildung 2.5 schematisch dargestellt:

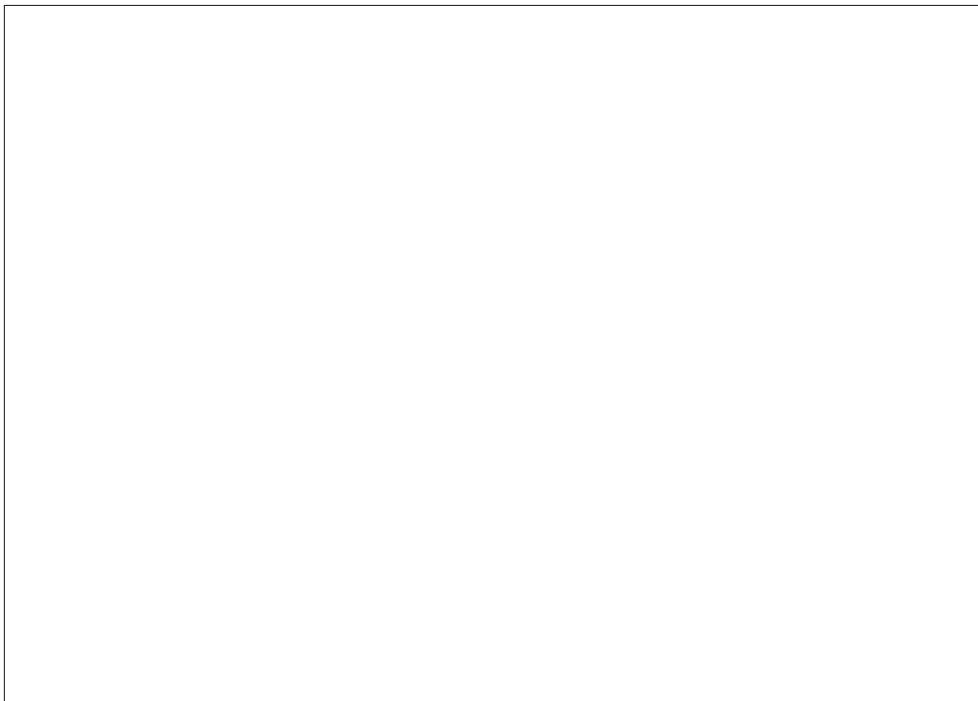


Abbildung 2.5: Funktionsweise des Verfahrens zur Erfassung von Trajektorien

Die Trajektorien werden im verwendeten Verfahren durch eine ID (Identifikationsnummer) und eine Trajektorienadresse (s. Abb. 2.5 - trackID, adress) beschrieben, an der die bereits passierten Positionen gespeichert werden.

Die Verschiebungsvektoren des optischen Flusses werden zunächst für jeden Eintrag nach Positionen durchsucht, an denen die Vektorlängen (Merkmalsgeschwindigkeiten) oberhalb einer festgelegten Geschwindigkeitsschwelle liegen. Wird eine solche Position gefunden, liegt dort ein sich deutlich bewegendes, interessantes Merkmal vor, dessen nähere Umgebung in der ID Map des Vorgängerframes nach bereits vorhandenen IDs durchsucht wird (vgl. Abb. 2.6). Wird keine andere ID in der Umgebung in der ID Map gefunden, wird eine neue ID für diese, neu erstellte, Trajektorie vergeben (vgl. Abbildung 2.5 ID Map - schwarz).

Zum Erweitern der vorhandenen Trajektorien wird für jede Pixelposition zusätzlich die ID Map des Vorgängerframes nach bereits vorhandenen IDs durchsucht.

Wird eine ID gefunden, wird umgekehrt die Umgebung Dieser in den Verschiebungsvektorlisten nach Positionen durchsucht, an denen die Merkmalsgeschwindigkeit oberhalb der Geschwindigkeitsschwelle liegt. Wird ein solches Merkmal gefunden, wird die ID für dieses Merkmal übernommen (vgl. Abbildung 2.5 ID Map - blau). Um die Positionen der Trajektorien zu aktualisieren, werden die erstellten/übernommenen IDs mit den Versätzen DX und DY des Verschiebungsvektors in eine neue ID Map eingetragen, die nach der Evaluierung eines Frames gespeichert wird, um im nächsten Frame zum Verknüpfen der Trajektorien (als ID Map des Vorgängerframes) zu dienen. Wird eine neue Trajektorie erstellt, wird dieser Versatz auf die Merkmalsposition addiert. Beim Erweitern einer bereits vorhandenen Trajektorie wird der Versatz auf die Position der ID addiert, um möglichst gerade Trajektorien zu erhalten. So beugt man instabilen Merkmalen vor, die ihre Detektionsposition an der Person ändern. Zugleich werden die IDs in einer Liste von Trajektorien (vgl. Abb. 2.5 - Track-List) zusammen mit allen Positionen eingetragen, die eine bestimmte Trajektorie bisher passiert hat. Verhält sich eine Trajektorie länger als die maximale Verweilzeit statisch, wird sie entfernt.

In nebenstehender Grafik erkennt man beispielhaft die Suche nach IDs von einem gefundenen Merkmal in der ID Map des Vorgängerframes in einer sog. 8er-Nachbarschaft. Hier ist keine ID vorhanden, weswegen eine neue ID für dieses Merkmal erstellt wird. Diese wird mit dem DX/DY-Versatz des Verschiebungsvektors in die neue ID Map eingetragen. Das rot markierte Pixel ist der Center Pixel, an dem ein sich bewegendes Merkmal vorliegt. Dieser Ablauf entspricht der Situation in Frame 1 in Abbildung 2.5 (schwarz).

Abbildung 2.6:
Suche nach
IDs

Systemarchitektur

In folgendem Kapitel wird die Implementierung des, in Kapitel 1 beschriebenen, Videoanalysesystems behandelt. Dabei wird zunächst in Abschnitt 3.1 der Aufbau des Systems und die Funktion der Bildauswertungssoftware als Teil der gesamten Systemarchitektur erklärt. Im Abschnitt 3.2 wird die Umsetzung der Eingangsgrößen (Bilder als Sensorsignale) in Messgrößen wie Personenzahl, -fluss oder -dichte in bestimmten Bildausschnitten beschrieben. Die Umsetzung erfolgt auf Basis der, in Kapitel 2 beschriebenen, Grundlagen. Weiterhin werden in Abschnitt 3.3 geplante Ablaufschritte zur Konfigurierung der Software beschrieben.

3.1 Beschreibung des Messsystems

Großveranstaltungen sind meist nicht mehr mit nur einer Kamera zu überblicken. Daher werden meist ganze Kameranetzwerke als Verbund mehrerer Kameras als Messsysteme eingesetzt. Ein möglicher Messaufbau wird in Abbildung 3.1 dargestellt.

Die Kameras sollten in ihrer Ausrichtung und ihren Abbildungseigenschaften statisch sein. Setzt man dies voraus, besitzen die Kameras fest definierte Anordnungen zueinander. Die Auflösung des Kamerabilds muss ausreichend sein, um einzelne Personen aufzulösen. Nachdem das System aufgestellt und eingerichtet ist, müssen Grenzwerte der Messgrößen für jede Kamera erlernt und die Konfigurierung, wie in Abschnitt 3.3 beschrieben, durchgeführt werden. Das Eingangsbild der Kameras wird in rechteckige Bildausschnitte eingeteilt, für die jeweils Statistiken und Messwerte generiert werden. Diese werden als Listen oder Tabellen übertragen. In der Überwachungszentrale (siehe Abbildung 3.1), in der die Datenströme der Kameras zusammengeführt werden, können, durch eine vorherige Kalibrierung, die Patch Positionen der Kamerabilder georegistriert auf Karten des Überwachungsgeländes übertragen werden.



Abbildung 3.1: Aufbau eines möglichen Videoanalysesystems

Damit können zudem die Messungen aus Bildausschnitten verschiedener Kameras zusammengeführt und auf eine einzelne Ortsposition bezogen werden. Geben die Kameras als Netzwerk an einer Ortsposition eine durchschnittliche Messgröße in einem kritischen Bereich aus, können automatisch Notrufe und Warnungen ausgesendet werden. Zusätzlich kann das Personal in der Überwachungszentrale eine generierte Karte des Überwachungsgeländes, die georegistrierte und gemittelte Messungen der versch. Kameras darstellt, analysieren. Hier werden die durchschnittlichen Messwerte der verschiedenen Kameras als Heat Map wahlweise mit angezeigten Messwerten dargestellt. Zusätzlich kann im Nachhinein nach der Veranstaltung das Bewegungsverhalten der Personengruppen im Detail analysiert werden, indem Überwachungsvideos gespeichert werden und in Ruhe das Verhalten der Messgrößen betrachtet wird.

Nachfolgend werden die Anforderungen des Systems an die Nutzer beschrieben:

Anforderungen des Systems:

- statische Rahmenbedingungen: statische Kameraperspektive
- Kalibrierung der Positionen/Perspektiven der Kameras (wird nicht im Rahmen dieser Arbeit behandelt)
- Konfigurierung mit, für die Szene sinnvollen, Evaluations-Parametern wie Such- oder Abstands-Radius
- Konfiguration mit intrinsischen und extrinsischen Kameraparametern zur Skalierung von Evaluations-/Filtereinstellungen und Extremwerten (z. B. in Form einer Patch-Karte)
- Trainieren von Grenzwerten, um relative Messwerte zu erhalten

3.2 Signalverarbeitung

Um innerhalb eines Bildes lokale, statistische Messergebnisse zu erhalten, wird das Eingangsbild zunächst in rechteckige Bildausschnitte (Patches) eingeteilt, deren Höhe und Breite im Konfigurationsdialog bestimmt werden können. Die Trajektorien werden dabei ausgewertet, um sie in die Bildausschnitte entsprechend einzugliedern und patchbezogene Messwerte zu erhalten. Dieses Verfahren stellt in dieser Arbeit die Grundlage für die Signalverarbeitung, also die Umsetzung der Eingangsgrößen in Messgrößen, dar. Die Laufvariable j ist in diesem Kapitel konsistent ein Index für die Frames einer Eingangsbildfolge und indiziert dabei den aktuellen Frame. Die Laufvariablen x, y indizieren eine Position im Eingangsbild. Diese müssen immer innerhalb des Wertebereichs $W(w, h)$ des Bildausschnitts liegen. Die nachfolgend beschriebenen Messgrößen werden in jedem neuen Frame j aktualisiert, d.h. neu berechnet und in sogenannte Patch Maps eingetragen.

3.2.1 Trajektorien

Die Umsetzung der Eingangsgrößen (-bilder) in Messgrößen erfolgt in dieser Arbeit auf Basis von erstellten Trajektorien (wie in Kapitel 2 Abschnitt 2.4 beschrieben).

Die Trajektorien können mit dem erstellten Softwaremodul als Linien gezeichnet werden. Die Erfassung solcher Trajektorien anhand einer künstlich generierten Bildsequenz erkennt man beispielhaft in Abbildung 3.2.

Abbildung 3.2: Erfassung und Darstellung von Trajektorien aus versch. Perspektiven [?] [?]

Hier erkennt man die automatische Erfassung von Trajektorien des entwickelten Verfahrens. Dabei werden 27 Frames einer künstlich generierten Bildsequenz analysiert, in der Personen, die anfangs in gleichem Abstand stehen, in alle Richtungen gegen die Wände drängen. Alle erfassten Koordinaten einer Trajektorie werden als zusammenhängende Linie in blau eingezeichnet. Nachfolgend erkennt man in Abbildung 3.3 die Erfassung und Darstellung von Trajektorien in weiteren Szenen des verwendeten Datensatzes „AGORASET”(siehe [?] [?]):

Abbildung 3.3: Erfassung und Darstellung von Trajektorien in diversen Szenarien [?] [?]

3.2.2 Personenzahl

Die Zahl der bewegten Personen N wird in ein Verhältnis mit der Summe an Trajektorienendpunkten (-spitzen) im Patch gesetzt. Um die Summe dieser zu erhalten, wird die Kardinalität (Mächtigkeit) der Menge S an Trajektorienspitzen (Trajektorien an ihrer Position im aktuellen Frame j), die sich innerhalb des betrachteten Bildausschnitts befinden, berechnet:

$$N = c_K \cdot |S| \quad [\text{personen}] \quad (3.1)$$

für alle verwendeten Positionen $P(x,y) \in S$ gilt: $x,y \in W(w,h)$

Hier bezeichnet $W(w,h)$ den zulässigen Wertebereich für die Bildkoordinaten x,y , falls diese sich im Bildausschnitt befinden. Dieser ist abhängig von der Breite w und Höhe h des Bildausschnitts.

Wendet man auf N einen zeitlichen gleitenden Mittelwert an, um Rauschen zu unterdrücken, erhält man die mittlere Personenzahl \bar{N}_j :

$$\bar{N}_j = \frac{1}{m} \sum_{i=0}^{m-1} N_{j-i} \quad [\text{personen}] \quad (3.2)$$

Um relative Messgrößen zu erhalten, kann, im Laufe einer Testsequenz, ein Maximalwert N_{max} für die (gemittelte) Personenzählung trainiert werden, der anschließend gespeichert wird. Die relative Personenzahl N_{rel} ergibt sich somit zu:

$$N_{rel} = \frac{\bar{N}_j}{N_{max}} \quad [\%] \quad (3.3)$$

mit $N_{rel} \in \{0..1\}$

Ist dieser Maximalwert von Hand in einem vollen Patch gezählt worden oder anderweitig genau bekannt, kann dieser absolut als N_{abs} im Konfigurationsdialog zusätzlich eingetragen werden. Dann kann aus der relativen Personenzahl im Bildausschnitt näherungsweise ein manuell definierbarer Normierungsterm (Kalibrierung des Zählsystems) N_{scal} in [personen] berechnet werden:

$$N_{scal} = N_{rel} \cdot N_{abs} \quad [\text{personen}] \quad (3.4)$$

Ist weiterhin die projizierte Fläche A des Bildausschnitts bekannt kann außerdem ein manuell definierbarer Normierungsterm für die Personendichte D_{scal} in $[\frac{\text{personen}}{m^2}]$ abgeleitet werden:

$$D_{scal} = \frac{N_{rel} \cdot N_{abs}}{A} \quad [\frac{\text{personen}}{m^2}] \quad (3.5)$$

3.2.3 Personenfluss

Der Personenfluss wird als der mittlere Nettozufluss an Trajektorien in einem Patch über die letzten p (Schrittweite) Frames definiert. Um in jedem Frame zu- oder abfließende Trajektorien zu erfassen, wird ein Fluss-Akkumulator F_{acc} eingerichtet, der den Beitrag +1 für eine in den Patch Eintretende und den Beitrag -1 für eine aus dem Patch austretende Trajektorie erhält. Dazu wird in jedem Frame die Differenz der Kardinalität der Menge S an Trajektorienspitzen und der Kardinalität der Menge K an Trajektorienknoten (Position der Trajektorien im vorherigen Frame $j - 1$) im Patch gebildet und über p Frames gemittelt.

$$F_{acc} = c_K \cdot (|S| - |K|) \quad \left[\frac{\text{personen}}{\text{frame}} \right] \quad (3.6)$$

für alle verwendeten Positionen $P(x, y) \in S \cup K$ gilt: $x, y \in W(w, h)$

Der mittlere Nettozufluss über die letzten p Frames beträgt somit:

$$F = \frac{1}{p} \sum_{i=0}^{p-1} F_{acc, j-i} \quad \left[\frac{\text{personen}}{\text{frame}} \right] \quad (3.7)$$

Wendet man auf F einen zeitlichen gleitenden Mittelwert über M Frames an, um Rauschen zu unterdrücken, erhält man den mittleren Personenfluss \bar{F}_j .

$$\bar{F}_j = \frac{1}{m} \sum_{i=0}^{m-1} F_{j-i} \quad \left[\frac{\text{personen}}{\text{frame}} \right] \quad (3.8)$$

Wird ein betragliches Maximum F_{max} dieses mittleren Personenflusses algorithmisch gelernt, kann ein relativer Fluss angegeben werden:

$$F_{rel} = \frac{\bar{F}_j}{F_{max}} \quad [\%] \quad (3.9)$$

mit $F_{rel} \in \{0..1\}$

Analog zur Personenzahl kann bei Bekanntheit von N_{abs} ein manuell definierbarer Normierungsterm für den Personenfluss erhalten werden:

$$\bar{F}_{scal} = \bar{F}_j \cdot \frac{N_{abs}}{N_{max}} \quad \left[\frac{\text{personen}}{\text{frame}} \right] \quad (3.10)$$

3.2.4 Dynamik

Zur Ableitung eines aussagekräftigen Dichtefaktors wird zusätzlich zur Personenzahl eine Kennzahl benötigt, die die vorherrschende Dynamik im Patch beschreibt. Dazu wird eine mittlere Merkmalsgeschwindigkeit im Patch eingeführt. Es wird erneut die Menge S an Trajektorien betrachtet, deren Spitzen (aktuelle Positionen) sich innerhalb des Bildausschnitts befinden. Es werden Vektorbeträge für jede dieser Trajektorien nach folgendem Prinzip berechnet, um die Bewegungen über die letzten q (Schrittweite) Frames zu erfassen. Im Fall, dass eine betrachtete Trajektorie noch keine q Frames existiert, wird die Trajektorie aussortiert:

$$\begin{aligned} \text{für } j - q \geq 0: \quad |\vec{b}_j|(x,y) &= \sqrt{[P_j(x) - P_{j-q}(x)]^2 + [P_j(y) - P_{j-q}(y)]^2} \\ \text{sonst:} \end{aligned} \quad (3.11)$$

für alle verwendeten Positionen $P(x,y) \in S$ gilt: $x,y \in W(w,h)$

Hier bezeichnet $P_j(x)$ die x-Position der Trajektorienspitze im aktuellen Frame j . $P_{j-q}(x)$ bezeichnet die x-Position der Trajektorienspitze vor q Frames.

Die mittlere Merkmalsgeschwindigkeit B_j eines Bildausschnitts erhält man nun durch Mittelung über alle Vektorlängen $|\vec{b}_j|(x,y)$ innerhalb des Patches:

$$B_j = \frac{1}{|Z|} \sum_{k=0} Z_k \quad \left[\frac{\text{px}}{q \cdot \text{frame}} \right] \quad (3.12)$$

für alle verwendeten Vektorlängen $|\vec{b}_j|(x,y) \in Z$ gilt: $x,y \in W(w,h)$.

Hier beschreibt Z die Menge der berechneten Vektorlängen $|\vec{b}_j|(x,y)$.

Um Rauschen zu unterdrücken, wird diese Kennzahl noch zeitlich gemittelt:

$$\bar{B}_j = \frac{1}{m} \sum_{i=0}^{m-1} B_{j-i} \quad \left[\frac{\text{px}}{q \cdot \text{frame}} \right] \quad (3.13)$$

Bei längerer Beobachtung der mittleren Merkmalsgeschwindigkeit in einer Testsequenz wird deutlich, dass sie in einem nicht vollgestauten (also unkritischen) Bildausschnitt immer oberhalb eines Schwellwertes liegt (nach Einschwingen des Systems). Dieser Schwellwert kann festgelegt werden, indem z. B. ein Bildausschnitt betrachtet wird, der stetig vollgestaut wird und damit eine abnehmende mittlere Merkmalsgeschwindigkeit aufweist. Der Benutzer liest diese Messgröße, kurz nachdem er die Situation als kritisch einstuft, ab und trägt sie als Schwellwert B_S ein. Der Schwellwert sollte, sobald kein Zu-/Abfluss in/aus dem Patch mehr existiert, schnell erreicht werden. Mit der Festlegung des Schwellwerts kann nun eine relative mittlere Merkmalsgeschwindigkeit B_{rel} berechnet werden:

$$B_{rel} = \frac{B_S}{\bar{B}_j} \quad [\%] \quad (3.14)$$

mit: $B_{rel} \in \{0..1\}$

Der Trägheitsfaktor B_{rel} ist eine Art Staufaktor, der hohe Werte für geringe Dynamiken im Patch annimmt. Eine kritische Situation, wie beispielsweise ein Stau, liegt aber nur dann wirklich vor, wenn sich viele Personen im Bildausschnitt befinden und eine geringe Dynamik in Kombination vorliegt.

3.2.5 Dichtefaktor

Schlussendlich kann aus den zuvor abgeleiteten Größen ein Dichtefaktor definiert werden, der nur auf hohe Personenzahlen und hohe Trägheit in Kombination reagiert und hohe Werte annimmt.

$$dFactor = N_{rel} \cdot B_{rel} \quad [\%] \quad (3.15)$$

mit: $dFactor \in \{0..1\}$

Der definierte Dichtefaktor funktioniert nur dann zuverlässig, wenn die Auflösung hoch genug ist, um Trajektorien von vereinzelt Personen zu erstellen. Treten Personengruppen hingegen nur noch als Anhäufung gleich heller Pixel auf, sind die Personen nicht separat erkennbar und die Personenzählung wird immer zu niedrig ausfallen, wobei gleiches für den Dichtefaktor gilt. Dieser Dichtefaktor macht weniger eine Aussage über eine Dichte in $[\frac{\text{Personen}}{m^2}]$ als über den aktuellen Grad der Gefährdung der Besucher im betrachteten Bildausschnitt. Denn für hohe Dichtefaktoren, die kritischen Situationen entsprechen, ist sowohl die Personenzahl als auch die Trägheit im Patch hoch.

3.2.6 Darstellung der Messgrößen - Aufbereitung

Das Eingangsbild wird vom implementierten Verfahren wahlweise mit angezeigten Messwerten und eingezeichneten Bildausschnitten manipuliert. Außerdem können Messgrößen für einen genaueren Überblick durch Heat Maps dargestellt werden. Heat Maps sind farbkodierte Karten (ähnlich einer Wärmeverteilungskarte). Dabei wird das Bild zunächst in ein Grauwertbild konvertiert, damit es selbst keine Farbinformation mehr enthält. Dann wird für jeden Pixel eine neue Farbe im HSV-Farbraum bestimmt. Im HSV (hue, saturation, value)-Farbmodell kann der berechnete Grauwert an der Pixelposition als Helligkeitswert value eingesetzt werden, während die Farbsättigung (saturation) auf den Maximalwert gesetzt wird. Der Farbton (hue) wird dann auf den Wertebereich einer Messgröße skaliert und je nach Betrag färbt sich der Bildausschnitt mit zunehmender Messgröße von cyanblau bis dunkelrot. Zur Skalierung der Messgröße ist ein Training mit einer Testsequenz erforderlich. Dabei werden Maximal-/Minimalwerte der Messgröße vom Verfahren erlernt.

Diese können in Konfigurationsdateien gespeichert und wiederverwendet werden. Hier wurde der gemessene Dichtefaktor zunächst auf einen Wertebereich von 0..1 und anschließend auf dem Farbkreis auf die Farben von Cyanblau bis Dunkelrot skaliert. Im HSV-Farbmodell entspricht dies einem zahlenmäßigen Farbton von 180-359. Formelmäßig kann man den Farbton hue so ausdrücken, falls X_{rel} allgemein eine relative Messgröße beschreibt:

$$\text{hue} = X_{rel} \cdot (359 - 180) + 180 \quad (3.16)$$

Ein Beispiel für die Darstellung einer Heat Map erkennt man in Abbildung 3.4:

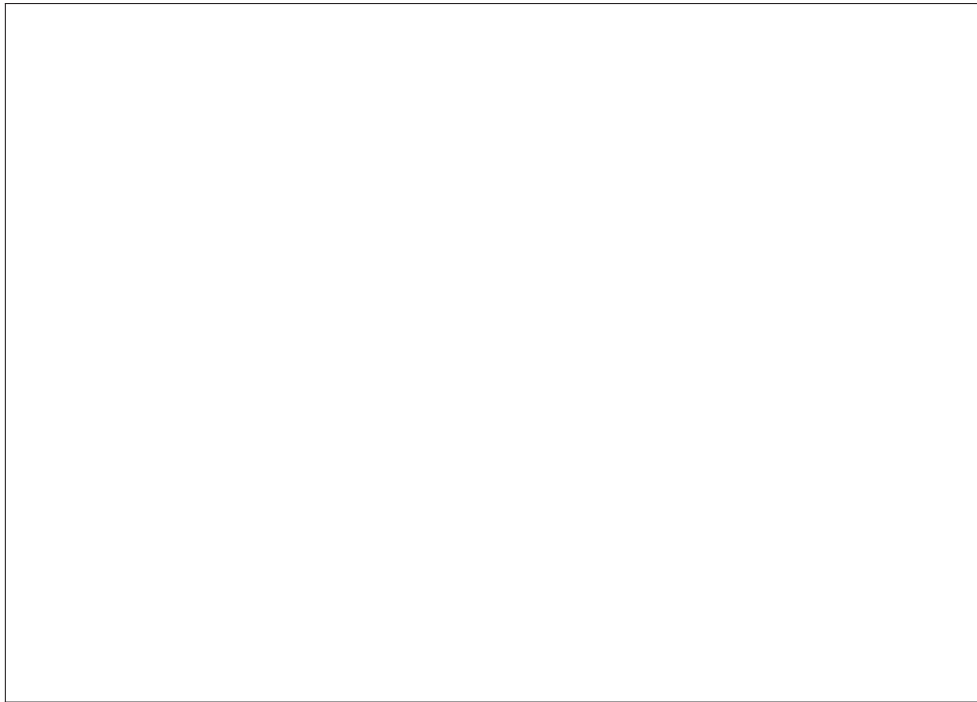


Abbildung 3.4: Darstellung des Dichtefaktors als Heat Map [?] [?]

Weil der Dichtefaktor aus der relativen Personenzählung und dem relativen Trägheitsfaktor abgeleitet wird, kann er nur für hohe Personenzahlen hohe Werte annehmen. Die maximale Personenzahl wird vom Verfahren in einer Lernphase statistisch geschätzt und tritt somit nur in einem voll befüllten Patch auf. Der abgeleitete Dichtefaktor beschreibt die Dichte der Personen im Bezug zur Fläche des Bildausschnitts. Personen, die zwar dicht aneinander stehen (siehe Abbildung 3.4 - lila), jedoch den Bildausschnitt nicht ausfüllen, erhalten keinen hohen Dichtefaktor. In einem solchen Fall würden aber in einem realen Szenario die Menschen sich instinktiv innerhalb des Bildausschnitts verteilen, weswegen eine Erkennung einer solchen Situation auch unerwünscht wäre.

Nur im Fall einer hohen Personenzahl und zusätzlich hoher Trägheit im Patch nimmt der Dichtefaktor hohe Werte an, wie in Abbildung 3.4 im dunkelrot gefärbten Bildausschnitt zu erkennen ist. Ist die Erkennung von kleineren, dicht gedrängten Personengruppen erwünscht, können die Bildausschnitte auch verkleinert werden, damit weniger Personen den Bildausschnitt bereits ausfüllen. Dies birgt allerdings die Gefahr für häufige Fehlalarme.

3.3 Konfigurierung

Der folgende Abschnitt befasst sich mit geplanten Ablaufschritten zur Konfigurierung des erstellten Softwaremoduls. Für optimale Ausnutzung des Funktionsumfangs müssen Grenzwerte geschätzt, Kameraparameter geladen, sowie Evaluations-, Filter- und Darstellungseinstellungen vorgenommen werden. Die Konfigurierung muss für jede Kamera/Sequenz einmal vorgenommen werden und kann dann gespeichert und wiederverwendet werden. Der Konfigurationsdialog des Softwaremoduls ist in Abb. 3.5 dargestellt

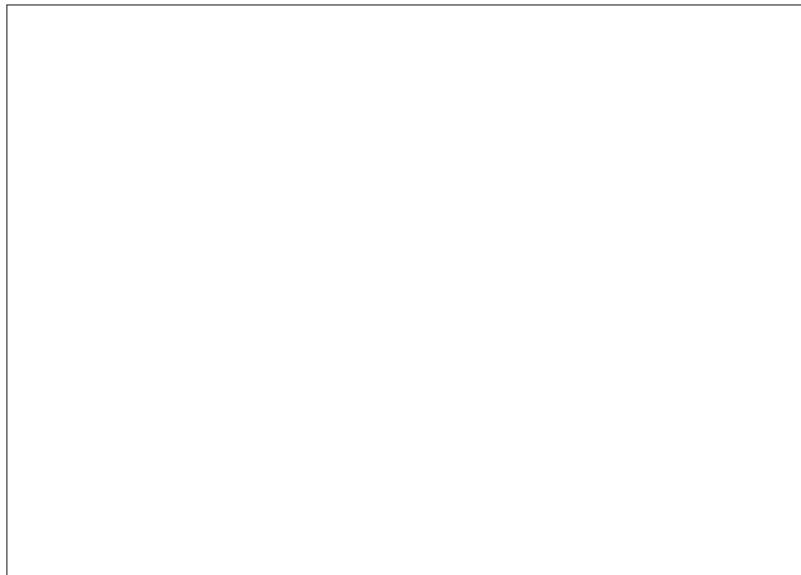


Abbildung 3.5: Konfigurationsdialog des Softwaremoduls

3.3.1 Patch-Map Dimension

Um die Patch Maps den Wünschen des Anwenders anzupassen, kann Höhe und Breite eines Bildausschnitts im Konfigurationsdialog gewählt werden. Anschließend muss das aktuelle Bild nach dem Betätigen eines Reset-Button neu geladen werden, um die Änderungen zu aktualisieren. Passen die gewählten Patches nicht gradzahlig ins Ausgangsbild, werden über den Rand überlappende Bildausschnitte am Rand erstellt. Wie bereits in Abschnitt 3.2 erwähnt, können die Patches wahlweise aufbereitet mit Messwerten oder Heat Maps eingezeichnet werden. Dieses aufbereitete Ausgangsbild wird vom Nutzer schließlich vereinfacht analysiert.

3.3.2 Geschwindigkeitsschwelle

Zur angepassten Unterscheidung von statischen Merkmalen zu Bewegten, wird eine konfigurierbare Geschwindigkeitsschwelle in $[\frac{px}{frame}]$ festgelegt. Zur Verwaltung (d.h. Erstellung und Fortführung) von Trajektorien werden nur Merkmale betrachtet, die oberhalb dieser festgelegten Geschwindigkeitsschwelle liegen. Statische Merkmale wie Ecken von Häusern oder Gegenständen werden somit weitgehend gefiltert, jedoch auch statische Personen. Eine geeignete Geschwindigkeitsschwelle erhält man empirisch, indem man zunächst den Suchradius vorübergehend auf 0 setzt, um ein Verknüpfen der Trajektorien weitgehend zu verhindern. So entstehen ständig neue Trajektorien an detektierten Ecken von Personen ohne unberechenbare Verknüpfungen. Anschließend vermindert man die Geschwindigkeitsschwelle stetig und lässt sich die Spitzen (aktuelle Positionen) der Trajektorien in der laufenden Bildsequenz einzeichnen. Die bewegten Personen im Bild sollten im Durchschnitt mindestens eine Trajektorie pro Person erhalten. Die höchste Geschwindigkeitsschwelle für die dies zutrifft, wird eingestellt.

3.3.3 Maximale Verweilzeit

Um zu erlauben, dass Trajektorien für eine gewisse Frame-Dauer statisch auf ihrer Position verharren, wird eine maximale Verweilzeit in [frame] gewählt. Dazu wird bestimmt, ob in der Umgebung einer Trajektorien-ID Merkmale oberhalb der Geschwindigkeitsschwelle liegen. Wenn ja, ist diese Trajektorie redundant, weil die Trajektorie fortgeführt wird. Wird kein solches Merkmal gefunden, kann die Trajektorie nicht fortgeführt werden und die Position der ID wird in einer Matrix (Stop-and-Go-Map) mit der Bildgröße inkrementiert, sofern der Wert die maximale Verweilzeit noch nicht erreicht hat. Sollte die Toleranz bereits erreicht sein, wird die Position in der Stop-And-Go-Map zurück auf 0 gesetzt und die Trajektorie entfernt.

3.3.4 Mittelungsdistanz

Wie in Abschnitt 3.2 behandelt, werden zeitliche Mittelungen im entwickelten Verfahren als gleitender Mittelwertfilter ausgeführt. Dazu kann im Konfigurationsdialog eine Mittelungsdistanz m festgelegt werden, die die Ordnung des Filters angibt, also über wieviele Frames der Mittelwertfilter ausgeführt wird. Meist eignet sich eine Mittelungsdistanz von 10 – 20 Frames, um Rauschen weitgehend zu unterdrücken.

3.3.5 Suchradius

Wie bereits in Kapitel 2 Abschnitt 2.4 behandelt, werden nur gefundene Merkmale, die oberhalb der Geschwindigkeitsschwelle liegen, behandelt. Wird ein Merkmal in der Umgebung einer ID in der ID Map gefunden, wird diese, bereits vorhandene, Trajektorie mit dem gefundenen Merkmal verknüpft und fortgeführt. Zur Suche nach Merkmalen in der Nachbarschaft ist die Definition eines Suchradius in [px] nötig, innerhalb dessen eine Suche nach Merkmalen/IDs vorgenommen wird. Dieser kann global im Konfigurationsdialog festgelegt werden. Einen geeigneten Suchradius für die Merkmale erhält man empirisch, indem man zunächst beim kleinsten Wert (1) beginnt und sich die Trajektorien, nach gewisser Dauer, in einem bestimmten Frame der Testsequenz einzeichnen lässt. Dies wiederholt man mit steigenden Radien, bis man durchgängige, lange Pfade erkennen kann. Der niedrigste Radius für den dies zutrifft, wird als Suchradius gewählt. Mit weiter steigenden Radien erscheinen die Pfade als eckiger/zackiger, weil die Merkmale großzügiger mit den ID's verknüpft werden. Damit ergibt sich eine höhere Toleranz für weit von der ID entfernte Merkmale, die möglicherweise schon zu einer anderen Person gehören und damit ein individuelles Tracking maßgeblich erschweren. Wechselt eine Trajektorie zwischen verschiedenen Personen, entstehen dabei Zacken. Gezeichnete Trajektorien werden in Abbildung 3.6 für verschiedene Suchradien dargestellt.

a) $r = 1$

b) $r = 2$

c) $r = 5$

Abbildung 3.6: Gezeichnete Trajektorien für versch. Suchradien

Man erkennt hier, dass für kleinere Radien (siehe a)) die Trajektorien fragmentiert wirken. Erhöht man den Radius (vgl. b)), funktioniert das Verknüpfen der Pfade zuverlässiger und man erkennt durchgängige Trajektorien. Wird der Radius wie in c) noch weiter erhöht, entstehen Zacken in den Pfaden, weil die Trajektorien großzügiger verknüpft werden und eine höhere Toleranz für Merkmale gesetzt wird, die von der ID weit entfernt liegen. Hier wurde der Radius 2 als Suchradius gewählt und gespeichert.

3.3.6 Abstands-Radius

Ebenso wie es zu vermeiden gilt, dass eine Trajektorie zwischen verschiedenen Personen wechselt, gilt es zu vermeiden, dass im zeitlichen Verlauf der Eingangssequenz im Durchschnitt unterschiedlich viele Trajektorien pro Person geführt werden. Dazu wird ein Abstands-Radius eingeführt, innerhalb dessen, nach der Evaluation eines Frames, die Nachbarschaft in der ID Map nach anderen IDs abgesucht wird. Wird eine ID gefunden, wird die länger bereits vorhandene Trajektorie behalten und die Kürzere entfernt, sofern ein Unterschied besteht. Damit unterdrückt man neue, kurze Trajektorien, die durch Rauschen entstehen und sorgt zudem dafür, dass die Trajektorien und damit auch die detektierten Personen zusammen eine große Fläche einnehmen müssen, um eine hohe Zahl an Trajektorien zu ermöglichen. Dies ist erwünscht, weil sonst beliebig viele Trajektorien an einer Person entstehen können, auch wenn diese alleine keine große Fläche einnimmt, was zu einer zu hohen Personenzählung führt. Eine große Zahl an Personen nimmt eine große Bildfläche ein und infolgedessen sollten auch nur in diesem Fall viele Trajektorien erlaubt sein. Durch Einführung des Abstands-Radius vermindert man Probleme, die durch gegenseitige Verdeckung der Personen entstehen. In weniger dichten Gebieten nimmt eine Person in der Regel eine größere Fläche ein, weil sie nicht durch andere Personen verdeckt wird. So können allerdings auch mehr Trajektorien an dieser Person entstehen, weil mehr Kanten dieser Person sichtbar sind. Nimmt die Personendichte an diesem Ort im Verlauf der Bildsequenz zu, wird diese Person mehr und mehr von Anderen verdeckt und es entstehen im Durchschnitt weniger Trajektorien pro Person, weswegen trotz der deutlich erhöhten Personenzahl, etwa die gleiche Zahl an Trajektorien und damit die gleiche Personenzählung wie zuvor vorliegt. Führt man einen Abstands-Radius ein, können nur bedingt viele Trajektorien an der unbedeckten Person entstehen und die Zahl an Trajektorien pro Person ist sowohl im Zustand geringer als auch hoher Personendichte in etwa gleich, wobei sich im Zustand hoher Dichte folglich, wie gewünscht, eine deutlich erhöhte Personenzählung ergibt.

3.3.7 Ortsbezogener Ausgleich

Bei Einführung dieses Abstands-Radius muss bedacht werden, dass die Ursprungsorte der Pixel im Eingangsbild, je nach Perspektive, unterschiedliche Abstände zum Kameraobjektiv aufweisen. Dieser Effekt wirkt sich umso stärker aus, je weiter man von der Vogelperspektive (Draufsichtperspektive) abweicht.

Demnach ist der Pixelabstand $d(x,y)$ in $[\frac{m}{px}]$ in entfernteren Bildregionen höher, weil solche Bildregionen kleiner skaliert abgebildet sind. Dies wird in Abbildung 3.7 verdeutlicht:

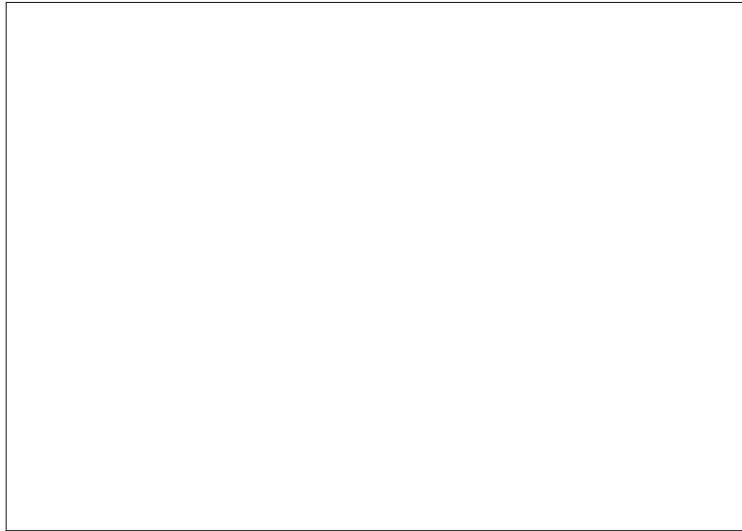


Abbildung 3.7: versch. Bildregionen mit unterschdl. Abständen zum Kameraobjektiv

Deshalb sind Menschen in weiter entfernten Bildregionen auch weniger Pixel hoch und voneinander entfernt. Sie nehmen damit, bei gleichbleibender Personenzahl, eine geringere Bildfläche ein, wobei weniger Trajektorien pro Person erfasst werden, weil die Kantenlänge in [px] verringert ist. Der Abstands-Radius sollte demnach in entfernteren Bildregionen sehr klein sein, um die ohnehin schon geringe Zahl an Trajektorien pro Person nicht weiter zu verringern. Dieser kleinste Abstands-Radius kann im Konfigurationsdialog festgelegt werden. Je näher sich das Objekt an der Kamera befindet, desto größer wird es abgebildet und der Abstands-Radius sollte mit einem entsprechenden (nichtlinearen) Skalierungsfaktor vergrößert werden. Infolgedessen sollte idealerweise eine Skalierungsfunktion in Form einer Bildmatrix vorliegen, in der der Pixelabstand (die Skalierung) in $\frac{m}{px}$ relativ zum größten Pixelabstand (zur kleinsten Abbildung) in $\frac{m}{px}$ im Eingangsbild für alle Bildkoordinaten x,y eingetragen ist. Mit Skalierungsfunktionen können Abstands-Radius oder andere Evaluations- und Filtereinstellungen sowie Grenzwerte skaliert werden. Eine solche Skalierungsfunktion kann leicht aus einer vollständig dokumentierten Kamerakalibrierung generiert werden. Alternativ kann auch eine Abstands-Radius Karte erlernt werden. Dazu muss der Abstands-Radius in px am unteren und oberen Bildrand als größter und kleinster Radius eingesetzt werden. Die Pixel zwischen oberer und unterer Bildzeile erhalten interpolierte Abstands-Radien zwischen den beiden Vorgegebenen. Dabei wird davon ausgegangen, dass die Entfernung vom Kameraobjektiv vom Unteren zum oberen Bildrand hin zunimmt, was oft in guter Näherung zutrifft. Wird keine Abstands-Radius-Karte vorgegeben wird der kleinste Abstands-Radius global verwendet. Für den kleinsten Abstands-Radius kann die Personenhöhe (in px) für den maximal zulässigen Abstand von Person zu Kamera eingesetzt werden (minimale Auflösung einer Person).

Alternativ kann auch eine geometrische Kalibrierung vorgenommen werden, um den Einfluss der unterschiedlichen Abstände der Bildregionen vom Kameraobjektiv zu verringern. Dazu werden bestimmte Bildregionen neu skaliert, so dass ursprünglich größer skalierte Gebiete, in denen also eine Person mit deutlich mehr Pixeln aufgelöst wird, eine geringere Auflösung erhalten, als Gebiete in denen die Personen ohnehin schon klein skaliert sind. Dabei verzerrt sich das Bild geometrisch. Da nun in allen Bildregionen eine Person mit näherungsweise gleich vielen Pixeln aufgelöst wird, werden nun Inhomogenitäten in der durchschnittlichen Zahl an Trajektorien pro Person vermieden. Dies wird in Abbildung 3.8 dargestellt:

Abbildung 3.8: Verfahren: geometrische Kalibrierung

3.3.8 Distanzschätzung

Sind Aperturwinkel ω , Tilt-Winkel T , sowie die Höhe h der Kamera bekannt, lassen sich näherungsweise die Abstände der Ursprungsorte von oberer und unterer Bildzeile vom Kameraobjektiv bestimmen. Dazu betrachtet man folgende Skizze in Abbildung 3.9:

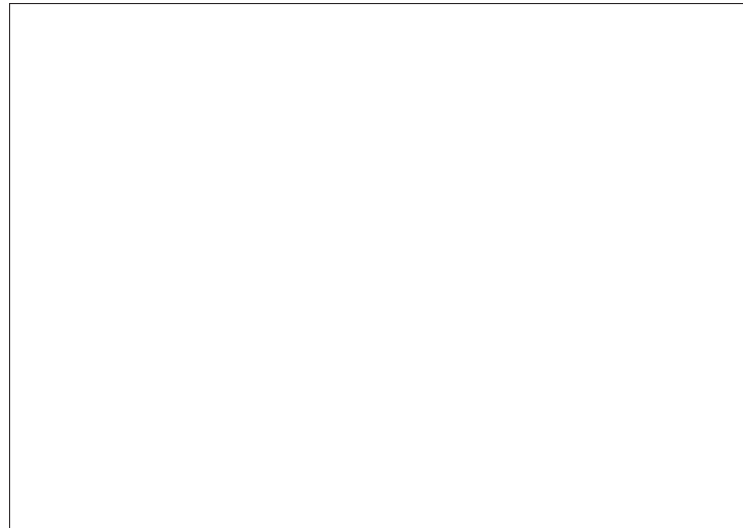


Abbildung 3.9: Distanzschätzung der Bildzeilen

Wendet man die trigonometrischen Gesetze an, wird hierbei deutlich:

$$\begin{aligned} d_1 &\approx \frac{h}{\sin(T + \frac{\omega}{2})} \\ d_2 &\approx \frac{h}{\sin(T - \frac{\omega}{2})} \end{aligned} \quad (3.17)$$

Die Abstände der Ursprungsorte von oberer und unterer Bildzeile vom Objektiv können mit diesen Formeln näherungsweise berechnet werden. Um eine Patch Map zu erhalten, die genäherte Distanzen der Bildauschnitte enthält, wird die Distanz zeilenweise zwischen oberer und unterer Bildzeile interpoliert und innerhalb der Patches gemittelt.

Liegt eine vollständig dokumentierte Kamerakalibrierung vor, ist dieses Verfahren redundant, weil aus der Kalibrierung der Abstand jedes Ursprungsorts im Eingangsbild berechnet werden kann. Dies ist jedoch für die Bilddaten, die für diese Arbeit vorliegen, nicht gegeben.

3.3.9 Filtereinstellungen

Leichte Kamerabewegungen lassen Trajektorien auch an Ecken von Gegenständen entstehen. Zusätzlich nehmen Wasser- und Windbewegungen (z. B. Blätter) Einfluss. Diese Bewegungen sind meist klein, können aber dennoch die gewählte Geschwindigkeitsschwelle betragsmäßig überschreiten. Erwünscht sind jedoch nur Trajektorien an Personen. Um solche fälschlicherweise entstandenen Trajektorien zu entfernen, können, vor der Umsetzung in Messgrößen, Filter auf die Trajektorien angewandt werden.

Nachfolgend werden die implementierten Filter aufgezählt:

- minimale Vektorlänge der Trajektorie über die letzten m Frames \Rightarrow kleine Kreisbewegungen werden gefiltert, exzentrische, gestreckte Bewegungen werden behalten
- minimale Eintragslänge der Trajektorie \Rightarrow kurze Trajektorien werden gefiltert, robuste, lange Trajektorien werden behalten
- maximale Eintragslänge der Trajektorie \Rightarrow besonders lange Trajektorien werden gefiltert

Evaluation

Das folgende Kapitel befasst sich mit der Auswertung des entwickelten Verfahrens. Dabei soll der erstellte Algorithmus erprobt und hinsichtlich seiner Zuverlässigkeit und Leistung untersucht werden. Um eine Erprobung vorzunehmen, müssen zunächst Referenzdaten durch manuelle Annotationen generiert werden. Eine solche manuelle Annotation, wird fachsprachlich als „Ground Truth“ (dt.: Grundwahrheit) bezeichnet. Das, im Rahmen dieser Arbeit entwickelte, Verfahren erhält die Bezeichnung „MFFT“ (multi frame feature tracking).

4.1 Testdaten

4.1.1 3D-Animation: AGORASET/Dispersion [?] [?]

Zum Testen der generierten Messungen des entwickelten Verfahrens wird eine vordefinierte, bekannte Szene verwendet, die in Form einer künstlich generierten Bildsequenz als Datensatz vorliegt. Die Szene zeigt initial gleichmäßig verteilte Personen in einer Halle aus der Vogelperspektive. Im Verlauf der Bildsequenz stäuben diese Personen von einem scheinbaren Ausgangspunkt in alle Richtungen auseinander. Nach einiger Zeit stauen sich die Personen an den Rändern der Halle. Die Sequenz besteht aus 110 Einzelbildern. In Abbildung 4.1 werden beispielhaft Ausschnitte der beschriebenen Sequenz gezeigt:

a) Bild Nr. 01

b) Bild Nr. 30

c) Bild Nr. 87

Abbildung 4.1: Ausschnitte der Testsequenz Dispersion [?] [?]

Das AGORASET ist ein Datensatz, der entwickelt wurde, um Algorithmen auf dem Gebiet der Analyse von Menschenmengen zu testen. Er besteht aus insgesamt 8 Szenen, die alle

künstlich generiert wurden, um Szenarien mit interessanten Personengruppeninformationen/-bewegungen zu erhalten.

4.1.2 Videosequenz: UCF_CrowdsDataset/Business-Getümmel [?]

Zusätzlich zur Evaluation der Messgrößen mittels einer künstlich generierten Bildsequenz soll die Evaluation mit realen Daten durchgeführt werden. Dazu wird eine Szene des UCF_CrowdsDataset (siehe [?]) verwendet, in der sich Personen ungerichtet im Alltagsstress auf einem Platz bewegen. Die Sequenz besteht aus 255 Einzelbildern. In Abbildung 4.2 werden Ausschnitte dieser Testsequenz gezeigt:

a) Bild Nr. 01

b) Bild Nr. 131

c) Bild Nr. 255

Abbildung 4.2: Ausschnitte der Testsequenz Business-Getümmel [?]

Anhand dieser realen Szene sollen Messgrößen wie Dichtefaktor, Personenfluss-/zahl zusätzlich evaluiert werden. Wegen stark varrierender Personenzahl und Dichte in verschiedenen Bildausschnitten ist diese Sequenz geeignet, um die Messgrößen hinsichtlich ihrer Sensitivität und ihrem Verhalten zu beurteilen. Die Szene liegt in einer maximalen Auflösung von 480x360 vor, weswegen diese Szene in besonders kurzer Laufzeit ausgewertet werden kann.

Das UCF_CrowdsDataset ist ein Datensatz mit aufgenommenen Bilddaten, der bewegte Menschenmengen in realen Szenarien zeigt. Er ist für die Erprobung von Algorithmen auf dem Gebiet der Analyse von Menschenmengen geeignet. Der Datensatz besteht insgesamt aus 30 Szenen, die aus öffentlichen Quellen (z. B. BBC) zusammengetragen wurden.

4.1.3 Videosequenz: UCF_CrowdsDataset/Marathon [?]

Um eine Testsequenz einzubeziehen, die extrem viele sowohl Bewegte als auch statische Personen enthält wird die Aufnahme eines Marathonlaufs verwendet. Die Sequenz liegt in Form einer Bildfolge mit einer Auflösung von 720x404 vor, weshalb dabei unter Ausführung des Verfahrens (MFFT) nur geringe Anforderungen an die Rechenleistung gestellt werden. In folgender Abbildung 4.3 werden beispielhaft Ausschnitte der Videosequenz des Marathonlaufs dargestellt.

a) Bild Nr. 01

b) Bild Nr. 31

c) Bild Nr. 62

Abbildung 4.3: Ausschnitte der Testsequenz Marathonlauf [?]]

Man erkennt hier deutlich, dass sich am Rande der Marathonstrecke stehende (statische) Personen befinden, während die Läufer sich bewegen. Daher wird diese Testsequenz dazu verwendet, die Trennung von statischen (HG) und nicht statischen (VG) Merkmalen zu evaluieren.

4.1.4 Videosequenz: Hamburger Hafenfest 2014

Um die Erfassung präziser Trajektorien in Personengruppen hoher Dichte zu evaluieren, wird das Szenario einer Massenveranstaltung verwendet. Dazu liegen Testsequenzen des Hamburger Hafenfests 2014 vor. Die Testdaten des Hamburger Hafenfests 2014 umfassen die Aufnahmen von einer in 80m Höhe montierten Kamera, die nach schräg links geneigt ist. Im Videobild ist hauptsächlich Wasser in Kanälen des Hamburger Hafens zu sehen. Während des Videos werden diese sporadisch von Schiffen befahren. 2 aufgehängte Brücken führen zur meist belebteren Hafenpromenade, die im Videobild teilweise sichtbar ist. Hier sind Buden und weiße Zelte zu erkennen, sowie Treppen und eine benachbarte Straße. Es liegen Aufnahmen von jeglichen Wetterlagen und Tageszeiten von dieser Kamera vor, um verschiedene Szenarien überprüfen zu können. In Abbildung 4.4 - 4.5 erkennt man beispielhaft Aufnahmen der verschiedenen Wetterlagen:

Die Aufnahmen liegen in einer maximalen Auflösung von 1920x1082 vor. Wegen hoher Rechenzeit beziehungsweise nicht verfügbarer Rechenleistung werden jedoch zum Testen skalierte Aufnahmen mit einer Auflösung von 960x540 verwendet. Moderne Computer können ein Eingangsbild einer solchen Bildfolge unter Ausführung des entwickelten Verfahrens MFFT in weniger als 50ms verarbeiten.

(a)

(b)

Abbildung 4.4: Testsequenz des Hamburger Hafenfests bei
(a) Nebel; (b) Regen

(c)

(d)

Abbildung 4.5: Testsequenz des Hamburger Hafenfests bei
(c) Sonne; (d) Sonnenuntergang

4.2 Testergebnisse

4.2.1 3D-Animation: AGORASET/Dispersion [?] [?]

Die Personen im Video sind mit einer Auflösung von 640x480 für eine Zählung ausreichend hoch aufgelöst. Zum Testen werden große Bildausschnitte mit einer Größe von 80x96 eingestellt und der Bildausschnitt gewählt, der den größten Wertebereich im Verlauf der Sequenz für die Personenzahl aufweist. Dieser Bildausschnitt ist am aussagekräftigsten für die Sensitivität und das Verhalten der Messgrößen (Personenzahl, -fluss, Dichtefaktor). Beim Testen werden nur relative Messgrößen betrachtet, weil, wie in Kapitel 3 Abschnitt 3.2 beschrieben, die Ableitung einer richtig skalierten absoluten Messgröße aus einer Relativen, durch Definition einer absoluten maximalen Personenzahl N_{abs} , jederzeit möglich ist. Diese wird entweder manuell oder durch theoretische Betrachtungen erhalten. Das entwickelte Verfahren wird, wie in Kapitel 3 Abschnitt 3.3 beschrieben, mit einem geeigneten Suchradius konfiguriert. Die maximale Verweilzeit für die Trajektorien wird auf eine längere Verweilzeit als die Dauer der Bildsequenz gesetzt, um stehenbleibende Personen mitzuzählen. Außerdem wird ein globaler Abstands-Radius eingestellt, der Probleme abmindert, die durch Verdeckungen entstehen. Die Schrittweiten p und q werden in diesem Abschnitt auf den Wert 10 festgelegt.

Personenzählung

Die Personen innerhalb des Bildausschnitts wurden für die 110 Frames der Testsequenz manuell gezählt und die Ergebnisse notiert. Um die Analogie der manuellen Zählung zum entwickelten Verfahren zu bewahren, werden die gezählten Personenzahlen mit einer Mittelungsdistanz von m Frames mit einem gleitenden Mittelwertfilter gemittelt. Die Mittelungsdistanz m wird hier auf 10 Frames festgelegt. Anschließend werden die Messungen des Verfahrens MFFT für die mittlere Personenzahl für den Verlauf der Testsequenz exportiert und gespeichert. Beide Ergebnisse werden über den zeitlichen Verlauf der Sequenz aufgetragen und gegenübergestellt, wie in Abbildung 4.6 dargestellt wird:



Abbildung 4.6: Graph: Verlauf der gemessenen/gezählten Personenzahl [?] [?]

Man erkennt hier, dass das Verhalten des erstellten Verfahrens (MFFT) mit dem Verhalten der Ground Truth korreliert. Im 1. Viertel der Sequenz erkennt man allerdings eine leicht erhöhte Zählung gegenüber der Ground Truth. Diese erhöhte Zählung tritt aufgrund von Verdeckungen auf, was auf dem Gebiet der Bildverarbeitung ein häufiges Problem darstellt. Dabei erhalten die Personen im Zustand geringer Personendichte (Beginn der Sequenz) deutlich mehr Trajektorien pro Person, weil eine frei stehende Person schlicht mehr Bildfläche einnimmt als eine Bedeckte, wodurch mehr Ecken (Harris Corners) sichtbar sind. Da das verwendete Verfahren die Spitzen der Trajektorien zählt, erhält man im Zustand geringer Personendichte eine zu hohe Personenzählung. Dieses Problem wird bereits vermindert, indem ein Abstands-Radius, wie in Kapitel 3 Abschnitt 3.3 beschrieben, eingestellt wird. Um es jedoch gänzlich aufzuheben, muss Aufwand betrieben werden, der den Umfang der Arbeit weit überschreiten würde. Es könnten beispielsweise Merkmalspunkte mit gleicher Bewegungsrichtung und -geschwindigkeit zusammengefasst werden (siehe Clustering), um pro Person jeweils nur eine Detektion zu erhalten.

Personenfluss

Das verwendete Verfahren zur Erfassung des Personenflusses stützt sich auf die Zählung von Trajektorienspitzen und -knoten (siehe Kapitel 3 Abschnitt 3.2) über die letzten p (Schrittweite) Frames. Die Schrittweite p wird hier auf 10 Frames gestgelegt. Dabei stellt man fest, dass der so erhaltene Personenfluss aufgrund der hohen Dynamik in der Szene stark fluktuiert und erneut gemittelt werden muss. Ab einer Mittelungsdistanz m von etwa 30 Frames erhält man eine robuste Schätzung. Dieses Verfahren kann in dieser Form nicht manuell annotiert werden, da der zeitliche Rahmen dieser Arbeit die Annotation ganzer Trajektorien, und dies wiederum für eine hohe Anzahl an Personen (Menschenmassen) nicht ermöglicht.

Als Alternative wird die Differenz der manuell gezählten aktuellen Personenzahl und der Personenzahl vor m (hier 30) Frames als absoluter Personenfluss über die letzten m Frames verwendet. Wird dieser durch die Mittelungsdistanz 30 geteilt, erhält man den mittleren Personenfluss der letzten m Frames, wie in Gleichung 4.1 dargestellt wird:

$$\bar{F} = \frac{N_j - N_{j-m}}{m} \quad (4.1)$$

Die so erhaltenen Messwerte werden über die Zahl der Frames aufgetragen und verglichen, wie in Abbildung 4.7 dargestellt wird:

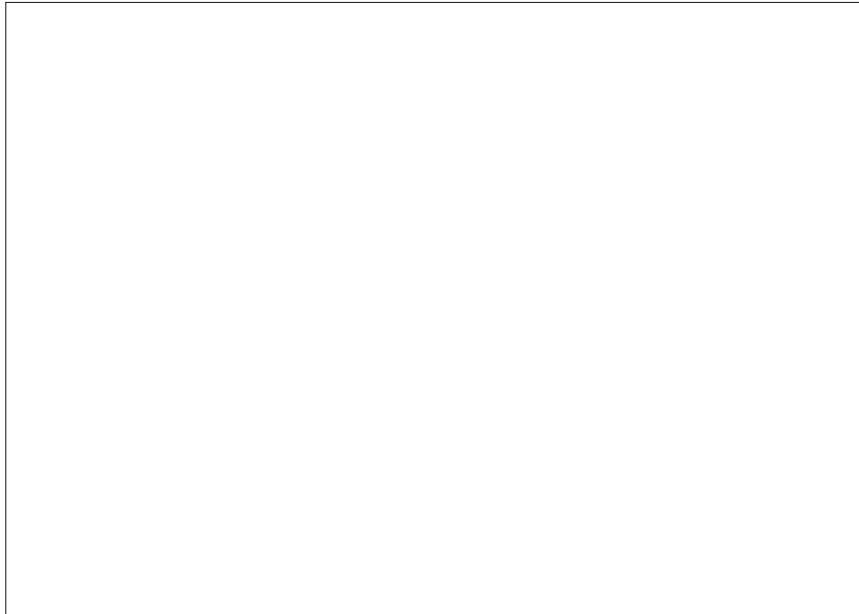


Abbildung 4.7: Graph: Verlauf des gemessenen/gezählten Personenflusses [?] [?]

Man erkennt, dass der vom entwickelten Verfahren abgeleitete Messwert für den Personenfluss zwar träge ist und eine Verzögerung aufweist, dem prinzipiellen Verhalten der Ground Truth aber folgt. Die so erhaltene Messgröße Personenfluss kann also als aussagekräftig betrachtet werden. Die Verzögerung tritt auf, weil sich das Verfahren zum Generieren der Ground Truth vom entwickelten Verfahren unterscheidet. Die Verzögerung um 10 Frames entsteht, weil die Berechnung des Personenfluss im entwickelten Verfahren (MFFT) zunächst über eine Schrittweite von p (hier 10) Frames berechnet wird. In manueller Form kann dieses Verfahren nicht ausgeführt werden, weswegen im Fall der Ground Truth diese Verzögerung ausbleibt.

Dichtefaktor

Der, in dieser Arbeit, abgeleitete Dichtefaktor stützt sich auf eine Kombination aus der Messung eines relativen Trägheitsfaktors (in Form einer relativen Merkmalsgeschwindigkeit im Patch) und einer Personenzahl (siehe Kapitel 3 Abschnitt 3.2).

Diese Messgrößen werden multipliziert um den Dichtefaktor zu erhalten. Da die manuelle Bestimmung der mittleren Merkmalsgeschwindigkeit der Personenbewegungen der letzten q (hier 10) Frames den Umfang der Arbeit überschreiten würde, wird für diese Messgröße keine Ground Truth durch Annotationen generiert. Der Verlauf des ausgegeben Dichtefaktors wird hier mit dem Verlauf der rel. Personenzahl verglichen, um zu zeigen, wie der abgeleitete Dichtefaktor funktioniert und sich von der reinen Messung einer Personenzahl unterscheidet. Die vom Verfahren (MFFT) generierten Messwerte werden exportiert und gespeichert. Die so erhaltenen Ergebnisse werden über den zeitlichen Verlauf der Testsequenz aufgetragen und betrachtet (siehe Abbildung 4.8):



Abbildung 4.8: Graph: Verlauf des Dichtefaktors und der Personenzahl (MFFT) [?] [?] (mit eingezeichneter Farbcodierung der verwendeten Heat Map)

Man erkennt hier, dass die Verläufe korrelieren, jedoch zu Beginn starke Abweichungen auftreten. Zu Beginn der Bildsequenz (Testbereich 1) hat die rel. Personenzahl bereits die Hälfte ihres Maximalwerts erreicht, weil sich bereits einige Personen im Bildausschnitt befinden. In diesem Bereich existiert jedoch noch eine hohe Dynamik (keine Staus) im Patch, weswegen kleine Messwerte für den relativen Trägheitsfaktor und damit auch für den Dichtefaktor generiert werden. Gegen Ende der Sequenz (Testbereich 2) beginnen die Personen im Bildausschnitt sich zu stauen, weswegen die Dynamik im Patch stark abnimmt. Damit werden immer größere Werte für den relativen Trägheitsfaktor und damit auch für den Dichtefaktor generiert, weswegen in diesem Bereich der Dichtefaktor stark ansteigt und mit der relativen Personenzahl in Übereinstimmung kommt. Ab hier nimmt der relative Trägheitsfaktor den Maximalwert 1,0 an, denn hier stimmen die Werte für Dichtefaktor und Personenzahl überein. Das Maximum des Dichtefaktors zu Beginn der Sequenz (Frame 10) entsteht durch den Einschwingvorgang der Messgröße relativer Trägheitsfaktor (vgl. Kapitel 3 Abschnitt 3.2).

Weil die mittlere Vektorlänge (B_j) über q Frames im Patch erst dann Messwerte generiert, wenn bereits q (hier 10) Frames vergangen sind und Diese noch zeitlich über m Frames gemittelt wird, dauert es weitere m (hier 10) Frames bis die mittlere Merkmalsgeschwindigkeit (\bar{B}_j) ihren hohen Betriebswert für die Dynamik annehmen kann (vgl. Fr. 20). Deswegen ist ihr Wert zu Beginn im Einschwingvorgang immer klein und erhöht sich dann auf den Betriebswert. Weil der relative Trägheitsfaktor (B_{rel}) als Schwellwert/Messwert ($\frac{\bar{B}_s}{\bar{B}_j}$) berechnet wird, wird man für kleine mittlere Merkmalsgeschwindigkeiten immer große relative Trägheitsfaktoren erhalten, was den Überschwinger zu Beginn der Sequenz erklärt.

Darstellung als Heat Map

Die Darstellung als Heat Map wird anhand des relativen Dichtefaktors in Abbildung 4.9 gezeigt. Dabei wird erneut die künstlich generierte Bildsequenz betrachtet, mit der die Messgrößen bereits evaluiert wurden. Der Verlauf des vom Verfahren (MFFT) ausgegebenen Dichtefaktors im Patch mit den Koordinaten (1|1) wird in Abbildung 4.8 - blau dargestellt. Zusätzlich wird auf der Ordinate die Farbcodierung der verwendeten Heat Map angegeben. Dieser Verlauf wird auf den Wertebereich, den die Heat Map für den Farbton (hue) vorsieht (180..359), skaliert. Hier werden exemplarisch Aufnahmen im Verlauf der Testsequenz Dispersion gezeigt, die mit Heat Maps aufbereitet werden. Die Heat Map stellt den Dichtefaktor dar. Kleine Werte für den Dichtefaktor ergeben die Farbtöne Cyan/Blau, hohe Werte ergeben die Farbtöne Lila/Rot. Wird dieser Dichtefaktor als Heat Map dargestellt, müssen, wie im Diagramm in Abbildung 4.8 dargestellt, zu Beginn der Testsequenz Cyan-/Blautöne (hue = 180 – 250) überwiegen. Gegen Ende (etwa ab Frame 70) der Testsequenz muss der Farbton in diesem Bildausschnitt ins Lila/Rote (hue = 300 – 359) wechseln.

a) Bild Nr. 10

b) Bild Nr. 70

c) Bild Nr. 100

Abbildung 4.9: Darstellung der Heat Map in der Testsequenz Dispersion [?] [?]

Zum Vergleich wird der Bildausschnitt mit den Koordinaten (1|1) betrachtet, da dieser Verlauf bereits in Abbildung 4.8 dargestellt ist. Man erkennt, dass in Frame 10 (vgl. a)) Cyan-/Blautöne überwiegen, sich die Messgröße zunächst einschwingt und dann im Arbeitsbetrieb befindet. Im Diagramm steigt der Dichtefaktor nun bis zum Ende der Testsequenz über die Farbtöne blau und lila (vgl. b)) auf seinen Maximalwert an, was ebenfalls zu erkennen ist. Wird, wie in c), der Maximalwert erreicht, wird der Bildausschnitt dunkelrot eingefärbt.

4.2.2 Videosequenz: UCF_CrowdsDataset/Business-Getümmel [?]

Die Testsequenz „Business-Getümmel“ wird verwendet, um erneut die Messgrößen Dichtefaktor, Personenfluss/-zahl mit realen Daten zu evaluieren. Die Personen sind mit einer Auflösung von 480x360 wegen der nahen Perspektive ausreichend aufgelöst. Für die Evaluation wird das Eingangsbild in nur einen Bildausschnitt mit einer Auflösung von 480x360 eingeteilt, weil ohnehin nur wenige Personen (maximal 60 auf einmal) in der Testsequenz zu sehen sind und ein möglichst hoher Wertebereich für die Messgrößen erzeugt werden soll. In der zuvor verwendeten künstlich generierten Bildsequenz waren bis zu 80 Personen in einem einzelnen Bildausschnitt zu sehen. Dieser Test mit realen Daten soll, in Analogie zum zuvor durchgeführten Test in Abschnitt 4.2.1, zur Evaluation des Verhaltens der Messgrößen ausgeführt werden. Erneut werden nur relative Messgrößen zum Testen verwendet, weil relative Messgrößen bei Bekanntheit eines Skalierungsfaktors jederzeit in richtig skalierte, absolute Messgrößen umgerechnet werden können (vgl. Abschnitt 4.2.1 und Kapitel 3 Abschnitt 3.3). Das Verfahren (MFFT) wird, wie in Kapitel 3 Abschnitt 3.3 beschrieben, konfiguriert und ein geeigneter Suchradius sowie die Geschwindigkeitsschwelle eingestellt. Die maximale Verweilzeit wird nun auf eine kurze Dauer gesetzt, da keine lang stehenden Personen in der Szene zu sehen sind. Die Schrittweiten p und q werden in diesem Abschnitt auf den Wert 10 festgelegt.

Personenzählung

Zur Durchführung der Tests müssen zunächst wieder Referenzdaten durch manuelle Annotationen generiert werden, die erneut durch manuelles Zählen der Personen in jedem Eingangsbild erhalten werden. Die Personen innerhalb des Bildausschnitts werden in 187 Frames manuell gezählt und die Ergebnisse notiert. Um erneut die Analogie zum verwendeten Verfahren (MFFT) zu bewahren, werden die gezählten Personenzahlen mit einem gleitenden Mittelwertfilter mit einer Mittelungsdistanz von m Frames gemittelt. Die Mittelungsdistanz m wird auf 10 Frames festgelegt. Anschließend werden die vom Verfahren ausgegebenen Messwerte für den Verlauf der Testsequenz exportiert und gespeichert. Die Ergebnisse werden in Abb. 4.10 dargestellt und betrachtet.

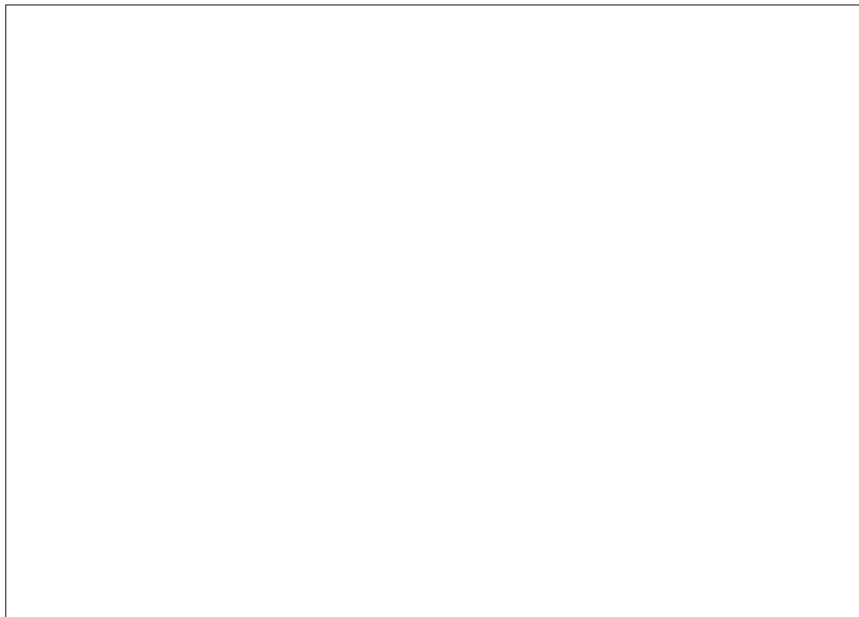


Abbildung 4.10: Graph: Verlauf der gemessenen/gezählten Personenzahl [?]]

Man erkennt hier die Auftragung der manuell Annotierten und der algorithmisch geschätzten Messwerte des entwickelten Verfahrens (MFFT). Die geschätzte Personenzahl folgt dem Trend der erstellten Ground Truth, ist jedoch rauschbehaftet. Vergleicht man die Graphen mit denen der Personenzählung aus Abschnitt 4.2 Abbildung 4.6, stellt man fest, dass die, zu Beginn auftretende, überhöhte Personenzählung hier ausbleibt. Die überhöhte Personenzählung tritt im Fall von Personenverdeckungen auf, die sich im Laufe einer Sequenz, in der die Personenzahl stetig zunimmt, ausbilden können. In der hier verwendeten Sequenz beträgt die Personenzählung initial nach dem Einschwingen über 60%. Hier nimmt also die Dichte nicht so drastisch (60%-100%) wie in der in Abbildung 4.6 verwendeten Sequenz zu, in der sie von 10%-100% zunimmt, weswegen eine Zunahme der Verdeckungen im Verlauf der Testsequenz hier sehr viel schwächer ausgeprägt ist. Die überhöhte Personenzählung zu Beginn der Sequenz bleibt also aus, weil bereits zu Beginn mehr als 60% der maximalen Personenzahl erreicht wird. Die Personen verdecken sich also demnach schon zu Beginn der Sequenz und verdecken sich im Verlauf der Sequenz nicht merklich zunehmend weiter.

Personenfluss

Wie bereits in Abschnitt 4.2.1 behandelt, kann das entwickelte Verfahren (MFFT) zur Erfassung des Personenflusses nicht manuell annotiert werden. Dieses Verfahren stützt sich auf das Zählen von Trajektorienspitzen und -knoten und es wäre sehr aufwendig für Menschen solche Trajektorien in einer Testsequenz zu erstellen, verwalten und schlussendlich auszuwerten. Daher wird erneut, wie bereits in Abschnitt 4.2.1, der Personenfluss über die letzten m Frames nach Gleichung 4.1 bestimmt. Die Schrittweite p wird auf 10 Frames und die Mittelungsdistanz m auf 30 Frames festgelegt. Der mittlere, relative Personenfluss wird über den Verlauf der Sequenz aufgetragen und in Abb. 4.11 dargestellt.



Abbildung 4.11: Graph: Verlauf des gemessenen/gezählten Personenflusses [?]]

Man erkennt, dass der Verlauf des mittleren Personenflusses erneut, wie schon mit den computergenerierten Daten, mit dem Verhalten der Ground Truth korreliert. Der Verlauf scheint jedoch um etwa 10 Frames verzögert. Die Verzögerung resultiert aus den unterschiedlichen verwendeten Verfahren. Die Verfahren erfassen beide einen Personenfluss, sind sich jedoch in ihrem Ansatz Einzelschätzungen zu glätten unterschiedlich. Diese Verzögerung tritt auf, weil die Ground Truth, wie in Gleichung 4.1 dargestellt, generiert wird. Im Verfahren jedoch, wird der Personenfluss zunächst über eine Schrittweite von p Frames berechnet und anschließend über m Frames gemittelt. Da dieses Verfahren in manueller Form nicht ausgeführt werden kann, wird der Personenfluss als Ground Truth direkt über m Frames erhalten, was die Verzögerung um p (hier 10) Frames erklärt.

Dichtefaktor

Wie bereits in Abschnitt 4.2.1 erwähnt, wird der abgeleitete Dichtefaktor aus der Multiplikation der Messgröße relativer Trägheitsfaktor B_{rel} mit der Messgröße relative Personenzählung erhalten. Da sich erneut, wie bereits in Abschnitt 4.2.1, das manuelle Vorgehen zur Bestimmung einer mittleren Merkmalsgeschwindigkeit den Umfang der Arbeit überschreiten würde, werden für den Dichtefaktor keine Referenzdaten durch manuelle Annotationen generiert. Der Dichtefaktor wird mit der, bereits evaluierten, rel. Personenzahl verglichen, um seine Funktion und Unterscheidung von der reinen Messung einer Personenzahl zu zeigen. Die geschätzten Messungen des Verfahrens für den Dichtefaktor werden mit denen für die relative Personenzahl in Abbildung 4.12 verglichen.

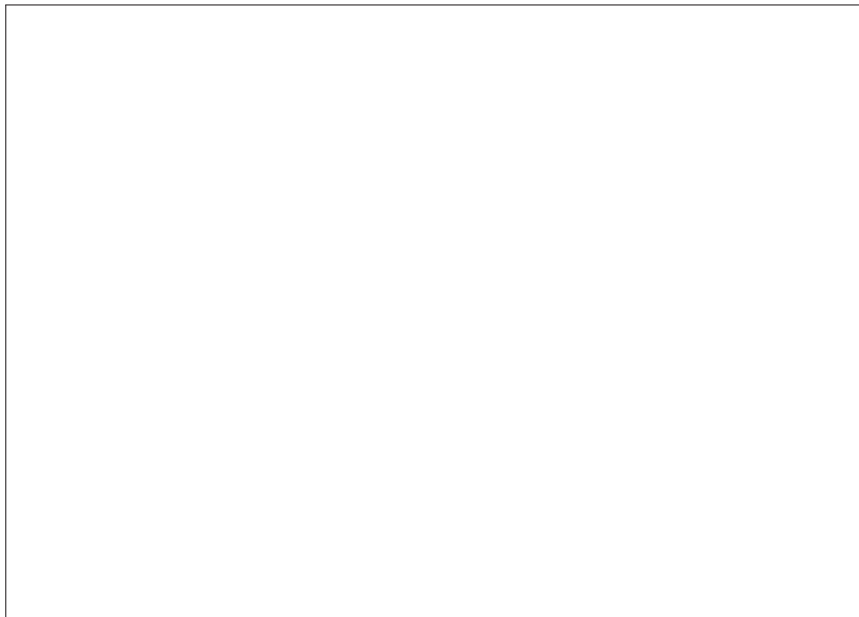


Abbildung 4.12: Graph: Verlauf des Dichtefaktors und der Personenzahl (MFFT) [?]]

Hier erkennt man, dass die Verläufe der beiden Messgrößen stark korrelieren. Die Abweichung zu Beginn der Sequenz, wie in Abschnitt 4.2.1 beschrieben, bleibt hier aus. Dies resultiert aus der hohen Dynamik, die im Verlauf dieser Testsequenz durchgängig vorliegt. Im Fall einer hohen Dynamik werden immer kleine Messwerte für den relativen Trägheitsfaktor generiert, wobei durch die Multiplikation gleiches für den Dichtefaktor gilt. Der Verlauf für den Dichtefaktor ist nach unten versetzt, weil die kleinen Werte des relativen Trägheitsfaktors multipliziert mit den hohen Werten für die relative Personenzahl eine nach unten versetzte Kurve ergeben.

Außerdem erkennt man den, aus Abbildung 4.8 bekannten, Überschwinger zu Beginn der Testsequenz. Dieser ergibt sich erneut durch den Einschwingvorgang der Messgröße relativer Trägheitsfaktor (vgl. Kapitel 3 Abschnitt 3.2). Weil zunächst q (hier 10) Frames vergehen müssen, bis die mittlere Vektorlänge der Bewegung über die letzten q Frames (B_j) Messwerte generiert und diese anschließend noch über m Frames gemittelt wird, dauert es weitere m (hier 10) Frames, bis die mittlere Merkmalsgeschwindigkeit (\bar{B}_j) ihren hohen Betriebswert annimmt. Diese Größe nimmt also zu Beginn im Einschwingvorgang kleine Werte an, weswegen der rel. Trägheitsfaktor (B_{rel}), der als $(\frac{B_s}{B_j})$ berechnet wird, zu Beginn immer hohe Werte annimmt. Weil aus dieser Größe der Dichtefaktor abgeleitet wird, nimmt dieser ebenfalls zu Beginn immer hohe Werte an und schwingt über. Nach dem Einschwingvorgang (etwa ab Frame 20) funktioniert diese Messgröße allerdings passabel und nimmt sinnvolle Werte an. Generell korreliert der Verlauf des abgeleiteten Dichtefaktors (MFFT) mit dem Verlauf der Ground Truth und kann als aussagekräftig betrachtet werden.

4.2.3 Videosequenz: UCF_CrowdsDataset/Marathon [?]

Die bereits beschriebene Videosequenz Marathon wird dazu verwendet, die Trennung von Bewegten und statischen Merkmalen zu evaluieren. Dafür wird die maximale Verweilzeit auf 0 Frame gestellt, um die Erfassung statischer Merkmale nicht zu erlauben. Das Verfahren wird wie gewohnt für die Szene konfiguriert. Anschließend werden über 30 Frames Trajektorien mit dem entwickelten Verfahren (MFFT) erfasst und gezeichnet. Die erfassten Trajektorien werden in Abbildung 4.13 dargestellt:

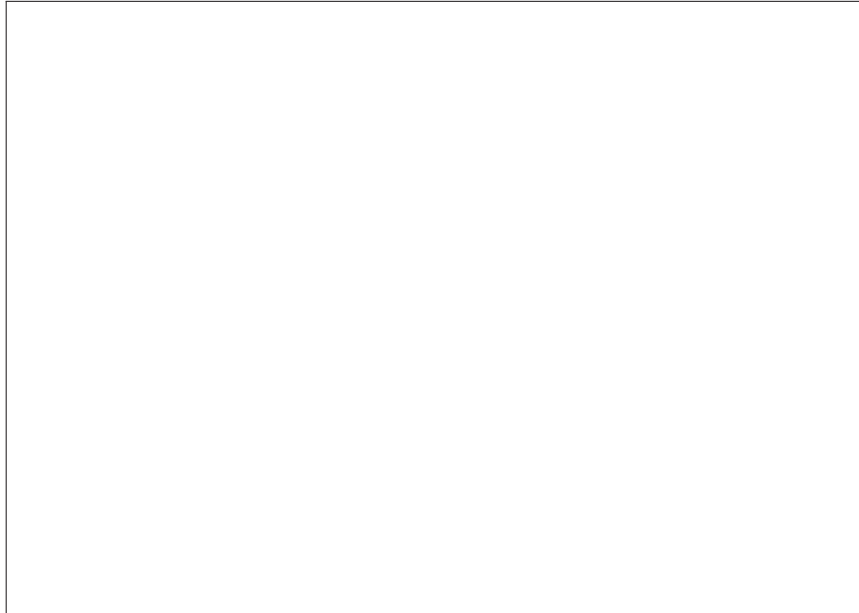


Abbildung 4.13: Darstellung von erfassten Trajektorien von Marathonläufern [?]

Man erkennt hier, dass nahezu ausschließlich in der bewegten Masse an Läufern Trajektorien erfasst werden, die als Gruppe oberhalb der eingestellten Geschwindigkeitsschwelle liegt. Statische Punkte wie Ecken an Häusern oder unbewegten Personen werden nicht erfasst. Die Trennung der Merkmale in Statische und nicht-Statische funktioniert also ausreichend und kann als plausibel angenommen werden.

4.2.4 Videosequenz: Hamburger Hafenfest 2014

Die zuvor beschriebene Videosequenz des Hamburger Hafenfests wird verwendet, um die Erfassung präziser Trajektorien in Personengruppen hoher Dichte zu evaluieren. Dabei werden niedrig aufgelöste Aufnahmen mit 960x540 verwendet, um zu zeigen, dass die Erfassung solcher Trajektorien auch mit relativ niedrig aufgelösten Daten zuverlässig funktioniert. Die Trajektorien werden erfasst und als blaue Linien im Eingangsbild in Abbildung 4.14 dargestellt:



Abbildung 4.14: Erfasste Trajektorien in der Testsequenz Hamburger Hafenfest

Es ist zu erkennen, dass mit dem erstellten Algorithmus selbst in Personengruppen hoher Dichte individuelle Trajektorien geführt werden können und somit die Ableitung von aussagekräftigen Messgrößen in derartigen Szenarien auf Massenveranstaltungen möglich ist. Außerdem erkennt man hier, dass die Fahnenbewegungen im Wind kleine Trajektorien entstehen lassen. Diese Trajektorien können gefiltert werden, indem man eine minimale Vektorlänge über die letzten M Frames vorgibt (vgl. Kapitel 3 Abschnitt 3.3). Da die Fahnenbewegungen Kreisbewegungen sind und keine exzentrische Orientierung haben, kann sich die Fahne nie über große Teile des Bildes bewegen, wie dies etwa Menschen tun. Damit wird sie nie eine hohe Vektorlänge über die letzten M Frames erreichen, weil sie immer wieder in etwa zu ihrem Ausgangspunkt zurückkehrt. Dabei werden zwar in Ausnahmefällen auch gewünschte Trajektorien gefiltert, jedoch können so Falschdetektionen (sog. False Positives) weitgehend verhindert werden.

Fazit

Das Ziel dieser Arbeit war es, ein algorithmisches Verfahren zur Analyse von Gruppierungs- und Bewegungsmustern zu entwickeln. Dieses sollte zunächst Trajektorien von Personen erfassen und analysieren. Aus den Trajektorien sollten weiterhin Messgrößen abgeleitet und dargestellt werden, mit denen eine möglichst genaue Analyse von Bewegungs- und Gruppierungsmustern möglich ist. Das entworfene Verfahren basiert auf dem Harris Corner Detector und dem sog. optischen Fluss (vgl. Kapitel 2). Es wurde im Rahmen eines Videoanalysesystems entwickelt, dessen Struktur in Kapitel 3 näher beschrieben wurde. Im Kapitel 4 wurde das entwickelte Verfahren mithilfe von Grundwahrheiten quantitativ bewertet. Dabei wurde festgestellt, dass die generierten relativen Messwerte mit dem Verhalten der manuell generierten Wahrheit (Ground Truth) korrelieren. Die generierten Messwerte können also als aussagekräftig betrachtet werden.

Probleme die zeitweilige Abweichungen verursachen sind z. B. Verdeckungen, die auftreten wenn im Verlauf der Sequenz die Personendichte stark variiert. Dieses Problem ist auf dem Gebiet der Bildverarbeitung bekannt und es existieren teilweise sehr aufwändige aber wirksame Lösungsansätze. Außerdem tritt oft zu Beginn der Sequenz ein Einschwingen bestimmter Messgrößen auf. Dies ist aber für die vorliegende Aufgabenstellung unkritisch, da das System für ein Langzeitmonitoring ausgelegt ist. Das vorliegende System wurde primär für die Erfassung bewegter Personen entwickelt und es konnte gezeigt werden, dass die Erfassung von Trajektorien in Personengruppen hoher Dichte, auch unter Verwendung von niedrig aufgelösten Bilddaten, zuverlässig funktioniert.

Als Ausblick könnte das Verfahren durch die Nutzung von intrinsischen und extrinsischen Kameraparametern, welche in der Regel durch eine Kalibrierung während der Einrichtungphase gewonnen werden, hinsichtlich Systemparametrisierung und Robustifizierung verbessert werden. So könnten Evaluations-/Filtereinstellungen und Grenzwerte über den Bildort x,y skaliert werden, um Einflüsse der unterschiedlich hohen Skalierung von Personen einzuschränken. Außerdem könnte versucht werden, dass Tracking der Personen durch die Wahl eines anderen Corner Detectors zu verbessern.

Möglicherweise existieren andere Eckendetektoren (z. B. SIFT, SURF, ORB uvm.), die in diesem Kontext eine höhere Erkennungsleistung liefern. Um stehende Personen mitzuzählen, ohne zugleich statische Objekte mit zu erfassen, könnte die maximale Verweilzeit nur auf Trajektorien angewandt werden, die bereits eine große Strecke zurückgelegt haben. So hätten spontan an Objektkanten entstandene Trajektorien keine Berechtigung weiter zu existieren. Zusätzlich könnte das Verfahren mit einem vereinfachten Konfigurationsablauf sowie einer Konfigurationsdokumentation implementiert werden.

Abschließend kann festgestellt werden, dass das entwickelte Verfahren vielversprechende Ergebnisse liefert und in einer weiteren Entwicklungsstufe unter realen Bedingungen evaluiert werden sollte. Insbesondere sollten Crowd Manager (Sicherheitspersonal bei Großveranstaltungen) die Praxistauglichkeit des Verfahrens bewerten und dadurch die Einsetzbarkeit in der Praxis bestätigen.

Alle URLs wurden zuletzt am 10.02.2016 geprüft.

Abbildungsverzeichnis

2.1	Betrachtung der Gradientenbeiträge verschiedener Patches	7
2.2	Verteilungen von Gradientenbeiträgen I_x, I_y [?]	8
2.3	Blockschaltbild: OCV_CornerTracker	10
2.4	optischer Fluss von detektierten Ecken an Marathonläufern [?]	12
2.5	Funktionsweise des Verfahrens zur Erfassung von Trajektorien	13
2.6	Suche nach IDs	15
3.1	Aufbau eines möglichen Videoanalysesystems	18
3.2	Erfassung und Darstellung von Trajektorien aus versch. Perspektiven [?] [?]	20
3.3	Erfassung und Darstellung von Trajektorien in diversen Szenarien [?] [?]	20
3.4	Darstellung des Dichtefaktors als Heat Map [?] [?]	25
3.5	Konfigurationsdialog des Softwaremoduls	26
3.6	Gezeichnete Trajektorien für versch. Suchradien	29
3.7	versch. Bildregionen mit unterschdl. Abständen zum Kameraobjektiv	31
3.8	Verfahren: geometrische Kalibrierung	32
3.9	Distanzschätzung der Bildzeilen	32
4.1	Ausschnitte der Testsequenz Dispersion [?] [?]	35
4.2	Ausschnitte der Testsequenz Business-Getümmel [?]	36
4.3	Ausschnitte der Testsequenz Marathonlauf [?]	37
4.4	Testsequenz des Hamburger Hafenfests bei (a) Nebel; (b) Regen	39
4.5	Testsequenz des Hamburger Hafenfests bei (c) Sonne; (d) Sonnenuntergang	40
4.6	Graph: Verlauf der gemessenen/gezählten Personenzahl [?] [?]	42
4.7	Graph: Verlauf des gemessenen/gezählten Personenflusses [?] [?]	43
4.8	Graph: Verlauf des Dichtefaktors und der Personenzahl (MFFT) [?] [?]	44
4.9	Darstellung der Heat Map in der Testsequenz Dispersion [?] [?]	45
4.10	Graph: Verlauf der gemessenen/gezählten Personenzahl [?]	47
4.11	Graph: Verlauf des gemessenen/gezählten Personenflusses [?]	48
4.12	Graph: Verlauf des Dichtefaktors und der Personenzahl (MFFT) [?]	49
4.13	Darstellung von erfassten Trajektorien von Marathonläufern [?]	50
4.14	Erfasste Trajektorien in der Testsequenz Hamburger Hafenfest	51